

Ein Rahmenwerk zur interaktiven Klassifikation hochauflösender optischer Satellitenbilder mittels graphenbasierter Bildmodellierung

vorgelegt von
Diplom-Vermessungsingenieur
Jens Kersten

von der Fakultät IV - Elektrotechnik und Informatik
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften
- Dr.-Ing. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. M. Opper
Berichter: Prof. Dr. O. Hellwich
Berichter: Prof. Dr. S. Hinz

Tag der wissenschaftlichen Aussprache: 20. September 2011

Berlin 2011
D 83

Zusammenfassung

Die vorliegende Arbeit behandelt die flächenhafte Informationsextraktion aus Satellitenbildern für die Bereitstellung von Informationsprodukten im Rahmen von Notfall- und Krisenkartierungen. Es wird ein interaktives Rahmenwerk zur Klassifikation sowie Vorklassifikation hochauflösender optischer Satellitenbilder konzipiert, prototypisch implementiert und empirisch evaluiert. Dabei wird zu Gunsten der Übertragbarkeit der Methode auf verschiedene Bildprodukte, Bildinhalte und thematische Fragestellungen keine pixelgenaue und vollautomatische, sondern vielmehr eine rasche interaktive Klassifikation von einfachen quadratischen Bildsegmenten angestrebt. Die Prozessierung erfolgt lediglich auf der Grundlage des Satellitenbildes selbst sowie den interaktiven Eingaben des Bildinterpreten, so dass eine Unabhängigkeit von zusätzlichen Daten gewährleistet ist.

Für eine wahrscheinlichkeitsbasierte Zuordnung der Segmente zu frei definierbaren thematischen Klassen wird ein hybrides Markov-Bildmodell vorgeschlagen, bei dem ein hierarchisches gerichtetes sowie ein ungerichtetes planares Markov-Zufallsfeld miteinander kombiniert wird. Das hierarchische Modell ermöglicht durch die Verwendung einer Quadtree-Baumstruktur eine rasche nicht iterative und exakte Inferenz der gesuchten Klassenzugehörigkeiten unter Einbezug von Bildinformation auf verschiedenen Skalen sowie unter Einbezug von kontextuellen Bedingungen zwischen den Klassen in benachbarten Skalen (hierarchischer Kontext). Das planare Modell beschränkt sich wiederum auf den Einbezug von räumlichem Kontext in der feinsten Ebene der hierarchischen Baumstruktur und kann kleinflächige Fehlklassifikationen sowie die durch das starre Quadtree-Modell hervorgerufenen Blockstrukturen im Inferenzergebnis verringern. Durch die Anwendung des hybriden Modells auf verschiedene Satellitenbilder kann gezeigt werden, dass durch Glättung kleinflächigen Fehlklassifikationen entgegen gewirkt wird. Um jedoch auch großflächigen Fehlklassifikationen, beispielsweise aufgrund von Mischsegmenten, entgegenzuwirken, werden in dieser Arbeit modifizierte hierarchische Inferenzmethoden vorgestellt und angewendet.

Die Schwerpunkte der Arbeit liegen bei der graphenbasierten Bildmodellierung und der Integration dieser Methodik in ein interaktives Rahmenwerk. Nach der Behandlung und Darstellung der statistischen Eigenschaften der Modelle sowie der Methoden der Parameterschätzung und Inferenz erfolgt daher die Darstellung des konzipierten Rahmenwerks sowie die Anwendung dieses auf sechs verschiedene IKONOS-Szenen. Insbesondere erfolgt eine ausführliche empirische Evaluierung, bei der für jeden Verarbeitungsschritt die jeweiligen Abhängigkeiten und die Sensitivität bezüglich der einfließenden Parameter und Trainingsdaten untersucht werden. Diese Evaluierung ist vor dem Hintergrund der stark interaktiv geführten Prozessierung sehr relevant und zeigt, dass die Güte der Ergebnisse vor allem von der Angabe einer repräsentativen Menge von Trainingsdaten abhängt. Durch die Evaluierung sowie den Vergleich der Ergebnisse des hybriden Modells mit denen zweier weiterer Verfahren wird der Mehrwert der Multiskalenprozessierung, der Modellierung kontextueller Relationen sowie insbesondere der Verwendung der hier behandelten modifizierten Inferenzmethoden demonstriert.

Abstract

In order to provide crisis mapping products especially during the *emergency phase*, i.e., immediately after disasters, a framework for classification and preclassification of high resolution optical satellite images is proposed, implemented and empirically evaluated. A general, fast and interactive approach for classification of different image products, image contents in the context of diverse thematic problems is proposed in this thesis. Since no pixel-level accuracy as well as full automatic processing is desired, simple square image segments are used. Image processing is done based on the satellite image itself as well as on input provided by an image analyst, which leads to an independency towards additional data sources.

Probability based labeling of image segments is done using a hybrid Markov-image model, which combines a hierarchical directed as well as a planar undirected Markov-random field. Due to the used quadtree-structure for the hierarchical model, fast, exact and non-iterative inference methods can be carried out incorporating image information from multiple scales as well as contextual constraints between classes in adjacent scales (hierarchical context). Due to the planar model, which is restricted to the incorporation of spatial contextual information in the finest quadtree-level, small-area missclassifications as well as „blocky“ structures can be reduced. In order to counteract large-area missclassifications, e.g. caused by mixtures of classes, a modified hierarchical inference method is proposed in this thesis.

The focus of this thesis is on graphical image modeling as well as on the formulation of a general framework using this methods. Therefore, based on the formulation of the statistical characteristics of the used models as well as the methods for parameter estimation and inference, the whole framework as well as its application is treated in detail. In particular an empirical analysis concerning the dependencies of the results with respect to the used model parameters as well as the training data for each processing step is carried out.

Particularly with regard to the interactive manner of processing, this evaluation is relevant. It can be reasoned, that the quality of the results is mainly dependent from the quality of the chosen training data. The results as well as a comparison of the proposed hybrid method with two other classification methods demonstrate the additional benefit of multiscale image processing as well as the proposed modified inference methods.

Danksagung

Diese Arbeit entstand im Rahmen des innerdeutschen Verbundprojekts DeSecure am Deutschen Zentrum für Luft- und Raumfahrt e.V. (DLR) in Oberpfaffenhofen. Die Betreuung der Arbeit seitens des DLR in der Abteilung Zivile Kriseninformation und GeoRisiken oblag Dr. Monika Gähler und Dr. Stefan Voigt. Bei ihnen möchte ich mich für die kontinuierliche Unterstützung durch wertvolle, kritische und lehrreiche Anregungen, ausführliche Diskussionen sowie ein stets und nicht nur auf Arbeitsebene offenes Ohr bedanken. All dies bot einen fruchtbaren Nährboden für meine spannende Tätigkeit in den vergangenen drei Jahren sowie den erfolgreichen Abschluss dieser Arbeit.

Von universitärer Seite wurde die Arbeit von Prof. Dr. Olaf Hellwich, Leiter des Fachgebiets *Computer Vision and Remote Sensing* des Instituts für Technische Informatik und Mikroelektronik der Technischen Universität Berlin betreut. Mit Prof. Dr. Olaf Hellwich fanden kontinuierlich über den gesamten Zeitraum der Dissertation überaus hilfreiche und wertvolle Gespräche und Diskussionen statt. Diese gaben stets Anlass, Sachverhalte aus anderen Perspektiven zu betrachten, Gedanken zu entwickeln und zu hinterfragen sowie Entscheidungen zu treffen und gaben damit entscheidende Impulse bezüglich sowohl der thematischen Ausrichtung als auch des zielführenden Fortschritts der Arbeit. Auch bei meinem Zweitbetreuer Prof. Dr. Stefan Hinz, Leiter des Instituts für Photogrammetrie und Fernerkundung am Karlsruher Institut für Technologie, möchte ich mich hiermit für die Unterstützung und Begleitung meiner Arbeit bedanken. Insbesondere das von ihm initiierte und jährlich organisierte Doktorandenkolloquium war eine sehr hilfreiche Plattform für den Austausch mit anderen Doktoranden.

In der ersten Hälfte der drei Jahre am DLR habe ich das Büro mit Dr. Sandro Martinis, in der zweiten Hälfte mit Steffen Bucher geteilt. Sandro danke ich hiermit für die hilfreichen Diskussionen zum Thema Markov-Bildmodellierung und Steffen für seine Bereitschaft und Fähigkeit, sich in verschiedenste Problemstellungen zu versetzen und seine Gedankengänge in konstruktive Schlussfolgerungen ergehen zu lassen.

Matthias Rupp und Clémence Dubois haben mich bei meiner Arbeit im Rahmen von Praktika sehr unterstützt, weshalb ich mich bei ihnen bedanken möchte. Beide haben maßgeblich bei der Bereitstellung von Referenzdaten sowie der Evaluierung von Methoden beigetragen. Clémence hat darüber hinaus bei dem Methodenvergleich sowie der Programmierung einzelner Softwaremodule mitgewirkt und sich intensiv mit dem Themengebiet der graphenbasierten Bildmodellierung auseinander gesetzt.

Alles bedeutet nichts, wenn man niemanden hat, der einem nahe steht, etwas in einem bewegt und mit dem man teilen kann. Die Menschen, die mir nahe stehen sind der Grund dafür, dass mein Leben genau so ist, wie es jetzt ist. Dafür bin ich sehr dankbar. Ich danke meinen Eltern für ihre Unterstützung und Anteilnahme auf meinem bisherigen Lebensweg. Bei meiner Freundin Nadja Kranz möchte ich mich für „das Leben, so wie es sein soll“, bedanken. Du lehrst mich jeden Tag aufs Neue, dass es die kleinen Dinge sind, die das Leben großartig machen. Es ist was es ist - und das ist wunderbar!

Inhaltsverzeichnis

Abbildungsverzeichnis	XIII
Tabellenverzeichnis	XV
1 Einleitung	1
1.1 Motivation	1
1.2 Rahmenbedingungen und Anforderungen an die Methodik	4
1.3 Methodische Einordnung	5
1.3.1 Bildsegmentierung und segmentbasierte Klassifikation	6
1.3.2 Kontextinformation und graphenbasierte Modelle	7
1.4 Zielsetzung der Arbeit	10
1.5 Gliederung der Arbeit	11
2 Kontextbasierte Klassifikation	13
2.1 Verfahren zur kontextbasierten Klassifikation	13
2.2 Graphenbasierte Bildmodellierung	15
2.2.1 Planare Modelle	16
2.2.2 Hierarchische Modelle	18
2.3 Diskussion	20
3 Wahrscheinlichkeitstheoretische graphenbasierte Bildmodellierung	23
3.1 Nichtkausale Modelle	23
3.1.1 Parameterschätzung	24
3.1.2 Inferenz	26
3.2 Hierarchische nichtkausale Modelle	26
3.2.1 Parameterschätzung	28
3.2.2 Inferenz	30
3.3 Kausale Modelle	30
3.3.1 Parameterschätzung	31
3.3.2 Inferenz	32
3.4 Hierarchische kausale Modelle	32
3.4.1 Parameterschätzung	33
3.4.2 Inferenz	34
3.5 Generativ versus diskriminativ	34
3.6 Diskussion	35
4 Hybrides Markov-Modell	39
4.1 Statistische Eigenschaften des hybriden Modells	39
4.2 Datenmodell	41
4.3 A priori Modell	41

4.4	Parameterschätzung	43
4.4.1	Parameterschätzung für das Datenmodell	43
4.4.2	Parameterschätzung für das a priori Modell	47
4.5	Inferenz	48
4.5.1	Hierarchische MAP-Inferenz	48
4.5.2	Hierarchische MPM-Inferenz	49
4.5.3	Modifizierte hierarchische Inferenz	50
4.6	Bewertung des hierarchischen Inferenzergebnisses	51
4.7	Kombination hierarchischer gerichteter und planarer ungerichteter Modelle . . .	52
4.8	Zusammenfassung	54
5	Integration des hybriden Modells in ein interaktives Rahmenwerk	55
5.1	Das Rahmenwerk	56
5.2	Anwendungsbeispiel	59
5.3	Zusammenfassung	62
6	Evaluierung und Vergleich der Methode	63
6.1	Beschreibung der Daten	64
6.2	Bildmerkmale	68
6.3	Identifikation relevanter Merkmale	69
6.4	Eigenschaften der Parameterschätzung für das Datenmodell	72
6.4.1	Szene (a): Ländliches Gebiet bei Dresden	73
6.4.2	Szene (b): Stadtgebiet von Dresden	75
6.4.3	Szene (d): Ländlicher Raum mit Industrie	78
6.4.4	Szene (f): Stadtrand von Dujiangyan, Sichuan, China	78
6.4.5	Zusammenfassung	79
6.5	Eigenschaften der Parameterschätzung für das a priori Modell	80
6.6	Bereitstellung von Trainingsdaten für das a priori Modell	83
6.7	Multiskalenprozessierung	85
6.8	Vergleich der Methode mit anderen gängigen Verfahren	91
6.8.1	Ergebnisse für Szene (a)	95
6.8.2	Ergebnisse für Szene (b)	96
6.8.3	Ergebnisse für Szene (d)	97
6.8.4	Ergebnisse für Szene (e)	98
6.8.5	Ergebnisse für Szene (f)	99
6.8.6	Zusammenfassung	99
6.9	Aussagekraft der Konfidenzkarte	101
6.10	Zusammenfassung	102
7	Zusammenfassung und Ausblick	105
8	Anhang: Verwendete Bildmerkmale	113
	Literaturverzeichnis	123

Abbildungsverzeichnis

1.1	Häufigkeit von Naturkatastrophen (links) und technischen Unfällen (rechts) zwischen 1975 und 2009. Quelle: http://www.emdat.be/ , <i>Emergency Events Database</i> (EM-DAT), Stand: Januar 2011	1
1.2	DLR-ZKI Übersichtskarte der Region Attica, Griechenland vom 25.08.2009. Landbedeckungsklassen: <i>Stadtgebiet</i> (rot), <i>Landwirtschaftliche Flächen</i> (beige), <i>Wald und Buschland</i> (grün), <i>Freier Boden / Dünne Vegetation</i> (grau) und <i>Wasser</i> (blau).	3
1.3	Zwei Bildelemente sind auf der Grundlage ihrer lokalen spektralen Information nicht voneinander unterscheidbar, obwohl sie eindeutig verschiedenen thematischen Klassen angehören (links). Nachbarschaftsinformation (die gelben Segmente kennzeichnen eine Nachbarschaft erster Ordnung) kann bei der Klassifikation zur Auflösung solcher Mehrdeutigkeiten beitragen (rechts).	8
1.4	Definition der Nachbarschaft der Bildelemente mittels Quadtree-Baumstruktur. Dargestellt ist ein Subgraph.	9
2.1	Schematische Darstellung eines Markov-Zufallsfeldes aus [Kumar, 2005]. Jedes Pixel i wird mit zwei Zufallsvariablen assoziiert. Die dunklen Knoten repräsentieren die Beobachtungen y_i , die hellen Knoten die gesuchten Klassenzugehörigkeiten x_i	16
3.1	Quadtree-Struktur für ein TCRF. Quelle: [Awasthi et al., 2007].	27
3.2	Einfache Struktur für ein Bayes'sches Netz. Quelle: [Ouyang et al., 2006].	30
3.3	Dyadische Darstellung der Graphenstruktur eines MSRF. Links: originale Struktur, Rechts: approximierte Struktur. Quelle: [Kumar, 2005].	32
4.1	Dyadische Darstellung des hybriden Modells. Jeder Knoten wird mit einer Klassenzugehörigkeit x (Attribut) sowie (optional) mit einem Merkmalsvektor y assoziiert. Die Merkmalsvektoren existieren in den Ebenen des Quadtrees, in denen auch Datenmodelle formuliert werden. Die Anzahl der Merkmale kann in jeder Ebene unterschiedlich sein. Die räumlich benachbarten Knoten im feinsten Level des Quadtrees sind ebenfalls mit einander verbunden (Nachbarschaft erster Ordnung).	40
4.2	Links: Ergebnis einer ML-Klassifikation. Mitte: Ergebnis mit $\chi^2_{\alpha,f}$ -Test mit $\alpha = 0.1$. Rechts: Ergebnis mit $\chi^2_{\alpha,f}$ -Test mit $\alpha = 0.3$. Ursprüngliches Bild und Referenz: Abbildung 5.2.	43
4.3	Schematischer Ablauf des <i>clusterings</i> nach [Ververidis und Kotropoulos, 2008].	45
4.4	Links: Eine Menge von zweidimensionalen Merkmalsvektoren. Rechts: Das dazu korrespondierende multivariate Testkriterium. Quelle: [Ververidis und Kotropoulos, 2008].	46

4.5	Ergebnisse der hierarchischen MPM-Inferenz. Links: die herkömmliche Inferenz (Gesamtklassifikationsgüte: 95.1%). Mitte: die modifizierte MPM-Inferenz (Gesamtklassifikationsgüte: 96.2%, $\alpha = 0.1$) verstärkt die Blockeffekte. Rechts: modifizierte Inferenz und anschließender ICM-Algorithmus (Gesamtklassifikationsgüte: 95.1%, $\beta = 1000$)	51
4.6	Entropiekarte berechnet aus den <i>a posteriori</i> Randdichten der MPM-Inferenz. .	52
4.7	Ergebnisse der kombinierten hierarchischen MPM- und planaren Inferenz (ICM-Algorithmus mit $\beta = 100$). Links: Ursprünglicher ICM-Algorithmus, Gesamtklassifikationsgüte: 96.1%. Rechts: Beschränkter ICM-Algorithmus mit $T_H = 0.001$, Gesamtklassifikationsgüte: 95.9%.	53
5.1	Schematische Darstellung der Klassifikation mit dem Rahmenwerk.	56
5.2	Links: IKONOS-Subset. Rechts: Referenz aus manueller Klassifikation.	59
5.3	Links: Dialog zur Erstellung des Quadrees. Rechts: Schematische Darstellung eines Teils der Baumstruktur.	59
5.4	Oben links: Dialogfenster für die Definition und Visualisierung der thematischen Klassen. Oben Rechts: Visualisierung der Trainingsdaten in Level 6. Unten: (v. l. n. r.): Trainingsdaten in Level 7 und 8.	60
5.5	2-d Visualisierung der Gauß-Modellierungen in den Merkmalsräumen der Level 6, 7 und 8 (v. l. n. r.).	61
5.6	Ergebnis der lokalen ML-Klassifikation mit $\alpha = 0.5$ für die drei Ebenen. Bei den schwarzen Bildelementen wurde die Nullhypothese verworfen.	61
5.7	Visualisierung der Übergangswahrscheinlichkeiten zwischen zwei Hierarchieebenen. Je dicker die Verbindung, desto höher die entsprechende Wahrscheinlichkeit. .	62
6.1	Subszenen der Größe 512×512 Pixel mit den entsprechenden Referenzklassifikationen aus visueller Interpretation.	66
6.2	Subszenen der Größe 1024×1024 Pixel mit den entsprechenden Referenzklassifikationen aus visueller Interpretation.	67
6.3	Ablauf des SFFS-Verfahrens nach [Somol et al., 1999].	71
6.4	Gesamtklassifikationsgüten für Szene (a): Ländliches Gebiet bei Dresden. Die Fehlerbalken stellen den Wert 0.5σ dar.	73
6.5	Konsumentengenauigkeiten der Klasse <i>Vegetation</i> (links) und Produzentengenauigkeiten der Klasse <i>Feld bewirtschaftet</i> (rechts) für Szene (a). Die Fehlerbalken stellen den Wert 0.5σ dar.	74
6.6	Links: Referenz. Rechts: Ergebnis einer ML-Klassifikation mit Fehlklassifikationen im rechten Bildbereich sowie vereinzelt in Bereichen von Klassenübergängen. .	75
6.7	Gesamtklassifikationsgüten für Szene (b): Stadtgebiet Dresden. Die Fehlerbalken stellen den Wert 0.3σ dar.	75
6.8	Konsumentengenauigkeiten der Klassen <i>Rest</i> und <i>Haus</i> . Die Fehlerbalken stellen den Wert 0.3σ dar.	76
6.9	Produzentengenauigkeiten der Klassen <i>Haus</i> und <i>Rest</i> . Die Fehlerbalken stellen den Wert 0.3σ dar.	77
6.10	Ergebnis einer ML-Klassifikation mit einer Gesamtklassifikationsgüte von 87.1 %. Links: Referenz. Rechts: Ergebnis mit teilweise falsch klassifizierten Straßenbereichen und Hausdächern.	77

6.11	Ergebnisse für Szene (d): Ländlicher Raum mit Industrie. Die Fehlerbalken stellen den Wert 0.5σ dar.	78
6.12	Gesamtklassifikationsgüten für Szene (f). Die Fehlerbalken stellen den Wert 0.5σ dar.	79
6.13	Links: Mittlere Produzentengenauigkeit der Klasse <i>Wasser</i> . Rechts: Mittlere Konsumentengenauigkeit der Klasse <i>Siedlung</i> . Die Fehlerbalken stellen den Wert 0.5σ dar.	79
6.14	Ergebnisse der Evaluierung des <i>a priori</i> Modells für die Szenen (a) bis (f). . . .	81
6.15	Trainingsdaten für die Schätzung der Parameter der <i>a priori</i> Modells in Abhängigkeit des Fehlerniveaus α	85
6.16	Visuelle Gegenüberstellung der Ergebnisse aus ML-Klassifikation und MPM-Inferenz für die Szenen (a), (b) und (c).	87
6.17	Visuelle Gegenüberstellung der Ergebnisse aus ML-Klassifikation und MPM-Inferenz für die Szenen (d), (e) und (f).	89
6.18	Konfidenzkarte und Ergebnisse der Methoden ML, SMAP, MPM-mICM, mMPM-mICM für Szene (a).	95
6.19	Konfidenzkarte und Ergebnisse der Methoden ML, SMAP, MPM-mICM, mMPM-mICM für Szene (b).	96
6.20	Konfidenzkarte und Ergebnisse der Methoden ML, SMAP, MPM-mICM, mMPM-mICM für Szene (d).	97
6.21	Konfidenzkarte und Ergebnisse der Methoden ML, SMAP, MPM-mICM, mMPM-mICM für Szene (e).	98
6.22	Konfidenzkarte und Ergebnisse der Methoden ML, SMAP, MPM-mICM, mMPM-mICM für Szene (f).	100

Tabellenverzeichnis

6.1	Metadaten der prozessierten Szenen.	65
6.2	Zusammenstellung der verwendeten Bildmerkmale.	68
6.3	Gesamtfehlerraten GF der beschränkten ML-Klassifikation sowie Gesamtklassifikationsgüten GK der hierarchischen MPM-Inferenz nach EM-Schätzung der <i>a priori</i> Modellparameter in Abhängigkeit vom Fehlniveau α	84
6.4	Gesamtklassifikationsgüten der hierarchischen MPM-Inferenz der Szenen (a), (b) und (c) bei verschiedenen Kombinationen von Skalen. Die grau hinterlegten Felder kennzeichnen das jeweils beste Ergebnis.	86
6.5	Gegenüberstellung der prozentualen Konsumenten- (UA) und Produzentengenauigkeiten (PA) aus ML-Klassifikation und der besten hierarchischen MPM-Inferenz für die Szenen (a), (b) und (c). Die grau hinterlegten Felder kennzeichnen eine Verbesserung durch die MPM-Inferenz.	86
6.6	Gesamtklassifikationsgüten der hierarchischen MPM-Inferenz der Szenen (d), (e) und (f) bei verschiedenen Kombinationen von Skalen. Die grau hinterlegten Felder kennzeichnen das jeweils beste Ergebnis.	88
6.7	Gegenüberstellung der prozentualen Konsumenten- (UA) und Produzentengenauigkeiten (PA) aus ML-Klassifikation und bester hierarchischer MPM-Inferenz für die Szenen (d), (e) und (f). Die grau hinterlegten Felder kennzeichnen eine Verbesserung durch die MPM-Inferenz.	88
6.8	Gegenüberstellung der Gesamtklassifikationsgüten aus ML-Klassifikation, SMAP-Inferenz und der vorgestellten Methode. Die grau hinterlegten Felder kennzeichnen das beste Ergebnis.	93
6.9	Gegenüberstellung der prozentualen Konsumenten- (UA) und Produzentengenauigkeiten (PA) aus ML-Klassifikation in ERDAS, SMAP-Inferenz in GRASS und bestem Ergebnis mit dem hybriden Modell für alle Szenen. Die grau hinterlegten Felder kennzeichnen jeweils Verbesserungen mit dem hybriden Modell (UA: hellgrau, PA: dunkelgrau).	94
6.10	Vergleich der Fehlerraten ER der MPM-Inferenz mit den prozentualen Anteilen der Segmente, die eine Entropie $H > T_H$ aufweisen und die zudem falsch klassifiziert wurden.	101

1 Einleitung

1.1 Motivation

Die Fernerkundung ermöglicht eine schnelle, effektive und großflächige Erfassung von Messwerten, welche Rückschlüsse über die aktuelle Situation auf der Erdoberfläche zulassen und stellt daher eine sehr wichtige und wertvolle Informationsquelle für Anwendungen im Bereich von Notfall- und Krisenkartierungen dar. Unmittelbar nach einer Naturkatastrophe oder einem technischen Unfall (die so genannte *emergency phase*) ist jede Information über die Situation vor und nach dem Ereignis von unschätzbarem Wert. Ein schneller und flächendeckender Überblick über die Situation hilft bei der Abschätzung des Ausmaßes sowie betroffene, gefährdete und stark hilfsbedürftige Bereiche zu detektieren, Hilfsmaßnahmen zu planen und Vorhersagen für weitere mögliche Ereignisse zu treffen. Eine sehr schnelle Auslieferung von aktuellen, hinreichend genauen und flächendeckenden Bildanalyseprodukten trägt signifikant zu der Erfassung großflächiger Notfallsituationen bei und stellt eine adäquate Entscheidungsgrundlage für involvierte (Hilfs-) Organisationen und Entscheidungsträger dar. Die globale Anzahl der

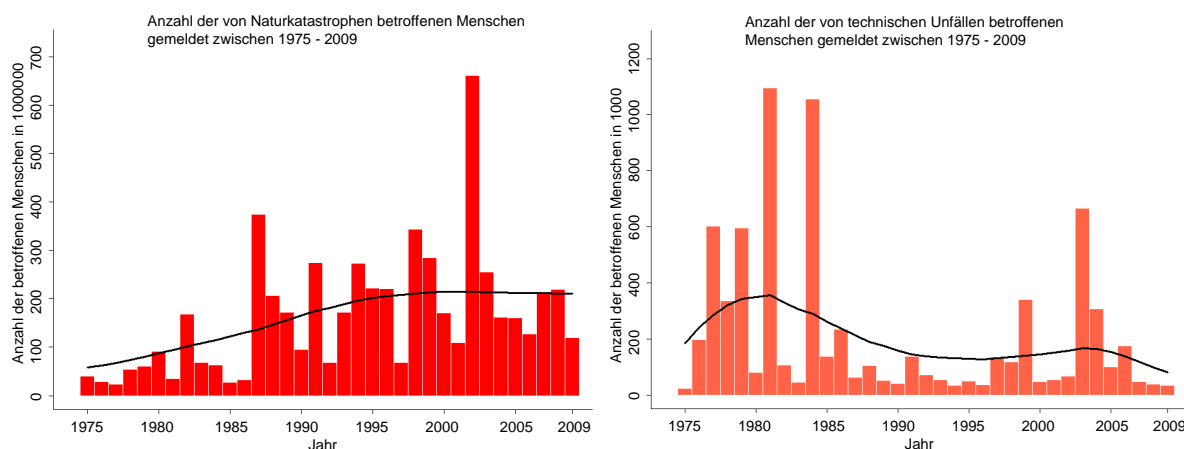


Abbildung 1.1: Häufigkeit von Naturkatastrophen (links) und technischen Unfällen (rechts) zwischen 1975 und 2009. Quelle: <http://www.emdat.be/>, *Emergency Events Database* (EM-DAT), Stand: Januar 2011

von technischen Unfällen und Naturkatastrophen betroffenen Menschen ist in den letzten Jahrzehnten auf einem sehr hohen beziehungsweise ansteigenden Niveau¹ (Abbildung 1.1). Parallel dazu ist in diesem Zeitraum auch die Verfügbarkeit von zivilen Erdbeobachtungssatelliten, die hochauflösende Satellitenbilder (räumliche Auflösung von 1 – 5 m) bereitstellen, gestiegen.

¹Für die Registrierung eines Desasters in die Datenbank EM-DAT ist mindestens eine der folgenden Fälle eingetreten: 10 oder mehr Menschen wurden getötet, 100 oder mehr Menschen waren betroffen, Notstand wurde ausgerufen oder internationale Hilfe wurde angefordert.

Aufgrund ihres hohen Informationsgehalts bilden hochauflösende Satellitenbilder sehr oft die Basis für die Bereitstellung verschiedenster Notfall- und Krisenkartierungsprodukte.

Die Fernerkundung beschäftigt sich intensiv mit der Konzipierung und Entwicklung von effektiven, schnellen und robusten Bildanalysemethoden, welche den implizit vorhandenen Informationsgehalt hochauflösender Satellitenbilder möglichst schnell verarbeiten können. Beispiele für kürzlich entwickelte Verfahren in diesem Bereich sind mit der unüberwachten Extraktion von flutinduzierten Änderungen der Rückstreuung in SAR-Daten (*synthetic aperture radar*) [Martinis et al., 2011] (Monitoring von Flutereignissen) sowie der Nahe-Echtzeit Bewertung von GIS-Objekten unter Verwendung multitemporaler Bilder [Frey und Butenuth, 2010] (Bewertung der Befahrbarkeit von Straßen in Krisengebieten) gegeben. Für den Einsatz solcher Bildanalysemethoden in operationellen Diensten, wie dem Zentrum für Satellitengestützte Kriseninformation² (ZKI), ist ein sehr hoher Automatisierungsgrad gewünscht. Die beim ZKI über die vergangenen Jahre gesammelten Erfahrungen zeigen, dass eine Automatisierung von Bildanalysemethoden aufgrund der vielfältigen und individuellen Fragestellungen, Datengrundlagen, Notfallsituationen sowie der Verwendung hochauflösender Satellitenbilder sehr schwierig ist und meist nur für spezielle Aufgabenstellungen und Rahmenbedingungen gelingt. Daher basiert die Analyse von Fernerkundungsdaten unter operationellen Rahmenbedingungen auch heute noch häufig auf visueller Interpretation [Voigt et al., 2007]. Dies begründet den Bedarf an neuen Bildanalysemethoden im Allgemeinen und insbesondere im Bereich von Notfall- und Krisenkartierungen.

Diese Arbeit entstand im Rahmen des deutschen Verbundprojekts DeSecure am Deutschen Zentrum für Luft- und Raumfahrt e.V. (DLR). Ziel dieses Projekts war es, die gesamte Prozesskette von Notfall- und Krisenkartierungen - von der Programmierung der Satelliten bis hin zur Weitergabe finalisierter Informationsprodukte an die Nutzer - zu optimieren. Vor diesem Hintergrund wird in dieser Arbeit ein allgemeines Konzept zur raschen Klassifikation und Vorklassifikation hochauflösender optischer Satellitenbilder konzipiert und evaluiert.

Unter einer Klassifikation ist dabei nicht die komplexe Aufgabe der pixelgenauen und automatischen Klassifikation oder gar Extraktion von höherwertigen krisenrelevanten Informationen, wie beispielsweise zerstörte Häuser, zu verstehen, sondern vielmehr die Zuordnung von quadratischen Bildsegmenten zu übergeordneten thematischen Klassen. Der Schwerpunkt der vorliegenden Arbeit liegt somit nicht bei einer bestimmten thematischen Fragestellung im Kontext von Krisenkartierungen, sondern bei der Bereitstellung einer Methode, die Kartierungsaktivitäten unterstützt. Aufgrund der oben genannten Schwierigkeit der Automatisierung und zu Gunsten einer Anwendung des Rahmenwerks bei unterschiedlichen Szenarien sowie auf unterschiedliche Bilddaten, wird dafür ein interaktiver Ansatz vorgeschlagen, bei dem das Bildverständnis eines Bildinterpreteten mit schnellen Methoden der Bildverarbeitung kombiniert wird. Dies soll eine Verringerung des Einsatzes visueller Komponenten im Sinne von manuellen Digitalisierungen und Nachbearbeitungsschritten erwirken, so dass Klassifikationsergebnisse für Notfallkartierungen mit mehr Schnelligkeit, Zuverlässigkeit und Reproduzierbarkeit aus hochaufgelösten Satellitenbildern abgeleitet werden können. Beispiele für Anwendungen dieser Methodik sind mit der Klassifikation einfacher Landbedeckungstypen oder mit der Vorklassifikation (z.B. die Ausweisung aller Bildbereiche, welche versiegelte Flächen reprä-

²Das ZKI ist ein Service des Deutschen Fernerkundungsdatenzentrums (DFD) im DLR. Die Aufgaben liegen bei der Beschaffung, Aufbereitung und Analyse von Satellitendaten bei Natur- und Umweltkatastrophen, für humanitäre Hilfsaktivitäten und für die zivile Sicherheit weltweit.

sentieren) auf der Grundlage beliebiger Satellitenbilder gegeben. Erstere könnte direkt für die Erstellung digitaler oder analoger Informationsprodukte (beispielsweise Karten) herangezogen werden (vergleiche Abbildung 1.2), eine schnelle Vorklassifikation einer Szene könnte hingegen als Zwischenergebnis für weitere automatisierte oder semiautomatische Analyseschritte, unter Verwendung spezieller Methoden der Informationsextraktion, dienen.

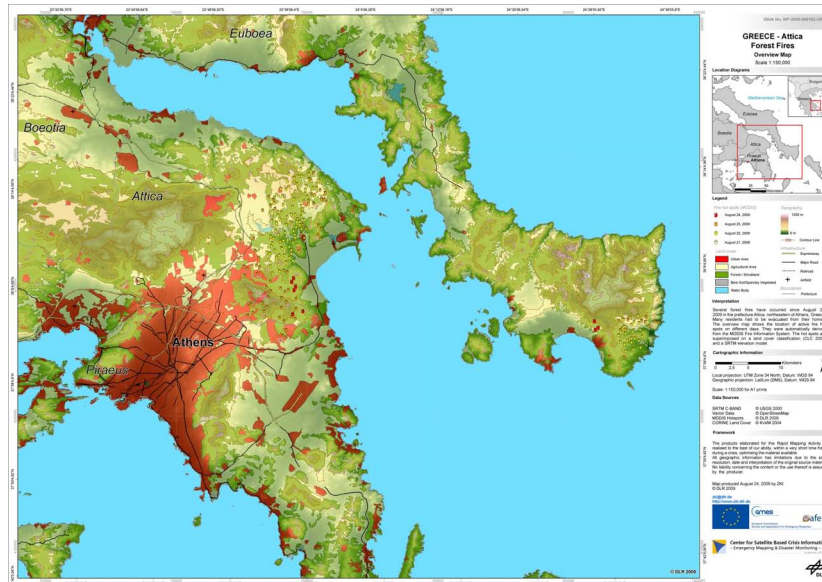


Abbildung 1.2: DLR-ZKI Übersichtskarte der Region Attica, Griechenland vom 25.08.2009. Landbedeckungsklassen: *Stadtgebiet* (rot), *Landwirtschaftliche Flächen* (beige), *Wald und Buschland* (grün), *Freier Boden / Dünne Vegetation* (grau) und *Wasser* (blau).

Die Verwendung von kontextueller Information ist letztlich unabdingbar für das Verständnis von Bildern [Pavlidis, 1986] und somit auch für ein leistungsfähiges Bildverarbeitungssystem [Li, 2009]. Die statistische graphenbasierte Bildmodellierung bietet einen konsistenten mathematischen Rahmen für die Integration von statistischen Modellen, die das Bildverhalten beschreiben, sowie von kontextuellen Bedingungen.

Aufgrund von Unsicherheiten, Rauschen und Mehrdeutigkeiten in den Bilddaten existiert üblicherweise keine vollkommen richtige und vollständige Lösung für bestimmte Aufgaben, beispielsweise der Klassifikation. Ansätze der statistischen graphenbasierten Modellierung basieren auf Prinzipien der Optimierung, so dass für die Interpretation von Bildern mit den oben genannten Effekten zwar möglicherweise inexakte, jedoch im Sinne der stochastischen Modellbildung optimale Lösungen gefunden werden können. Des Weiteren bietet die graphenbasierte Bildmodellierung das Fundament für eine Multiskalenprozessierung [Gidas, 1989]. Insbesondere vor dem Hintergrund einer Minimierung von manuellen Schritten, wie beispielsweise der Definition von Parametern nach dem *trial and error* Prinzip, wird in der vorliegenden Dissertation ein Ansatz der graphenbasierten Bildmodellierung vorgestellt.

Im Folgenden werden einige grundlegende, in dieser Arbeit verwendete Begriffe näher erläutert: Ein optisches Satellitenbild ist das Abbild eines begrenzten Bereichs der Erdoberfläche mit einer definierten Aufnahmegeometrie sowie einer bestimmten räumlichen und spektralen

Auflösung. Die realen semantischen Objekte, wie beispielsweise Häuser oder Bäume, sowie die realen Landbedeckungstypen, wie beispielsweise *Ackerland* oder *Wasserflächen*, weisen der Aufnahmegeometrie und Auflösung entsprechende Charakteristika im Satellitenbild auf. Dies ermöglicht prinzipiell die Detektion von semantischen Objekten und thematischen Klassen im Bild. Ein semantisches Bildobjekt in einem Satellitenbild entspricht dem Abbild eines realen semantischen Objekts und setzt sich in der Regel aus mehreren zusammenhängenden Bildpixeln zusammen. Im Gegensatz zu einem Bildobjekt ist ein Bildsegment oder Segment lediglich eine Gruppierung von mehreren Pixeln, der *a priori* keine semantische Bedeutung zugesprochen wird. Die Gruppierung kann beispielsweise durch die Betrachtung der spektralen Werte benachbarter Pixel oder auch willkürlich erfolgen. Äquivalent zu einem Pixel, kann auch ein Segment ein Bildobjekt, Bildobjektteil, ein Element einer Landbedeckung oder aber auch eine Mischung dieser darstellen. Die Klassifikation ist die Zuordnung von Pixeln oder Segmenten zu semantischen Klassen und wird in der vorliegenden Dissertation behandelt. Das bedeutet, dass hier semantische Objekte nicht durch explizite Modellierung detektiert werden.

1.2 Rahmenbedingungen und Anforderungen an die Methodik

Im Rahmen von Notfall- und Krisenkartierungen geht die Anwendung von Methoden der Bildverarbeitung mit verschiedenen zu berücksichtigenden Anforderungen, sowohl der Nutzer der Analyseergebnisse als auch der Nutzer der Methoden selbst sowie mit verschiedenen Rahmenbedingungen einher. Damit sind entscheidende Richtlinien bei der Konzipierung von Methoden in diesem Anwendungsbereich und daher auch für diese Arbeit gegeben.

Laufzeit: Eine schnelle Prozessierung von Satellitendaten ist besonders in der *emergency phase* essentiell. In der aktuellen Version des operationellen Dienstes ZKI ist die Auslieferungszeit von Produkten, welche das Ausmaß von Ereignissen sowie Schäden ausweisen, auf maximal acht Stunden nach Erhalt der Satellitendaten limitiert [Schneiderhan et al., 2010]. In dieser Zeit muss, neben der Informationsextraktion, auch die Vorprozessierung der Daten (sofern erforderlich) sowie die Erstellung und Auslieferung von standardisierten Karten und digitalen Produkten erfolgen. Die Prozessierung von Satellitenbildern sollte demnach so schnell wie möglich erfolgen.

Datengrundlage: Die zur Verfügung stehenden Daten und Satellitenbilder sowie die Charakteristika dieser können je nach Szenario sehr stark variieren. Häufig treten Naturkatastrophen und technische Unfälle an Orten auf, für die neben den ad-hoc akquirierten Satellitenbildern nur sehr ungenaue, nicht aktuelle oder gar keine Zusatzdaten, wie digitale Höhenmodelle, Vektordaten (beispielsweise von Straßen und Häusern) oder Satellitenbilder vor dem Ereignis zur Verfügung stehen. Eine rasche Prozessierung aktueller Satellitendaten muss auch in solchen Fällen möglich sein.

Übertragbare Anwendbarkeit: Ständig wechselnde Szenarien, thematische Fragestellungen und Datengrundlagen führen zu der Forderung nach einer allgemein formulierten Methodik. Die gewünschte schnelle Prozessierungszeit impliziert dabei eine schnelle und möglichst automatisierte Anpassung etwaiger Modellparameter. Die Anwendbarkeit derselben Methode zur Bildklassifikation in unterschiedlichen Notfallsituationen würde die Prozessierung enorm vereinfachen.

Räumliche und thematische Anforderung: Die gewünschte räumliche und thematische Auflösung des Klassifikationsergebnisses ist abhängig vom jeweiligen Szenario, der benötigten Information und den zugrunde liegenden Daten. Vor allem in der *emergency phase* kommt Daten und Informationsprodukten mit einer niedrigen Auflösung und Genauigkeit eine oft wichtigere Rolle als „vorqualifizierten“ Daten zu [Annoni et al., 2010], da diese in der Regel schneller verfügbar sind. Die im ZKI gesammelten Erfahrungen bestätigen diese Aussage.

Bewertung eines Bildanalyseprodukts: Die Ergebnisse von Bildanalysen im Rahmen von Notfall- und Krisenkartierungen stellen oft eine Entscheidungsgrundlage für (Hilfs-) Organisationen und Entscheidungsträger dar und können mitunter helfen Menschenleben zu retten. Aus diesem Grund besteht die Forderung nach einer adäquaten Bewertung der Analyseergebnisse. Damit kann der Bildinterpret auf möglicherweise falsche oder unsichere Ergebnisse hingewiesen sowie dem jeweiligen Nutzer detaillierte Informationen über die Aussagekraft und Sicherheit des bereitgestellten Informationsprodukts geliefert werden.

Anforderungen der Nutzer der Bildverarbeitungsmethoden: Ein wichtiger Punkt bei der Konzipierung der Klassifikationsmethode in der vorliegenden Dissertation ist die Forderung nach einer Minimierung der manuellen Arbeitsschritte. Diese bezieht sich sowohl auf die Teil- oder Nachbearbeitung durch visuelle Interpretation, als auch auf etwaige interaktive Eingaben, wie beispielsweise die manuelle Definition und Anpassung von sensiblen Modellparametern. Andererseits ist für die oben beschriebene übertragbare Anwendbarkeit ein hohes Maß an Flexibilität gefordert, was mit verschiedenen Möglichkeiten zur Visualisierung und Kontrolle von Zwischen- und Endergebnissen bei möglichst wenig erforderlichem Expertenwissen einher geht. Des Weiteren sollen reproduzierbare Ergebnisse generiert werden, welche durch geeignete Indikatoren qualitativ bewertet werden können.

Zusammenfassend können die oben formulierten Bedingungen und Anforderungen vor dem Hintergrund der zuvor beschriebenen aktuellen Situation im operationellen Kartierungsdienst ZKI als Forderung nach einem allgemeinen Ansatz für eine schnelle und interaktive Klassifikation und Vorklassifikation hochauflösender optischer Satellitenbilder angesehen werden. Der Begriff Interaktivität bezieht sich dabei auf prozessierungsbedingte und steuernde Eingriffe des Bildinterpreten. Eine zentrale Rahmenbedingung in dieser Arbeit ist mit der sehr begrenzten Datengrundlage gegeben. Neben dem Satellitenbild selbst stellt demnach das Bildverständnis des Bildinterpreten die einzige zusätzlich verfügbare, jedoch reichhaltige Informationsquelle für den Analyseprozess dar.

1.3 Methodische Einordnung

Bei der Klassifikation von optischen Satellitenbildern wird üblicherweise zugrunde gelegt, dass jede semantische Klasse eigene charakteristische Ausprägungen in den jeweils gemessenen spektralen Bereichen aufweist. Bayes'sche statistische Ansätze fassen sowohl die Parameter des jeweiligen Modells als auch die beobachteten Größen als Zufallsvariablen auf, wodurch Unsicherheiten und natürliche Variationen innerhalb der semantischen Klassen im Merkmalsraum mathematisch dargestellt werden können. Rauschen führt zu einer Verschmierung der ideellen spektralen Messwerte und somit auch zu einer Erhöhung der Varianz der Klassen im Merkmalsraum. Dieses Rauschen kann in klassenabhängiges und klassenunabhängiges Rauschen

unterteilt werden. Während ersteres auf natürliche Variationen in den spektralen Bändern innerhalb der Klassen zurückzuführen ist, wird letzteres durch externe Einflussfaktoren, wie atmosphärische Störungen sowie Variationen der Sensorcharakteristika, hervorgerufen [Sharma und Sarkar, 1998]. Bei hochauflösenden Fernerkundungsdaten tritt oft eine erhöhte Varianz der Pixelstatistik innerhalb von semantischen Klassen auf, da ein Pixel hier keine Mischung der spektralen Signatur unterschiedlicher semantischer Objekte oder Klassen, sondern einen Teil eines solchen Objektes oder einer solchen Klasse repräsentiert [Blaschke und Strobl, 2001], [Hese, 2001]. Aus diesem Grund wurden in den letzten Jahren vermehrt segmentbasierte Klassifikationsverfahren vorgeschlagen und erfolgreich in verschiedensten Fragestellungen angewendet.

1.3.1 Bildsegmentierung und segmentbasierte Klassifikation

Der Informationsgehalt hochauflösender Satellitenbilder nähert sich heutzutage immer mehr dem von Luftbildern an. Eine teilautomatisierte oder automatisierte Auswertung respektive Klassifikation solcher Bilder mittels herkömmlichen, auf statistischen Analysen beruhenden und auf Pixelebene arbeitenden Verfahren führt in vielen Fällen zu keinem zufriedenstellenden Ergebnis [Schneider et al., 2000]. Mit steigender räumlicher Auflösung übernimmt die Textur sowie die Objektgeometrie bei der Interpretation eine immer größere Rolle, wobei parallel dazu die Aussagekraft der spektralen Information abnimmt [Schneider et al., 2000]. Wie oben beschrieben, nimmt die Varianz der Pixelstatistik der semantischen Klassen mit steigender räumlicher Auflösung zu und äußert sich bei der Anwendung klassischer pixelbasierter Verfahren im so genannten „Salz-und-Pfeffer“ Effekt. Ein weiterer wichtiger, jedoch oft ignorierte Aspekt bezüglich der pixelbasierten Klassifizierung von hochauflösenden Fernerkundungsdaten ist, dass sich das Signal eines Areals der Erdoberfläche, welches durch ein Pixel repräsentiert wird, zu einem substantiellen Anteil mit den Signalen der benachbarten Pixel korreliert ist [Townshend et al., 2000].

Segmentbasierte Ansätze, wobei unter dem Begriff Segment Pixelgruppen, die ein definiertes Kriterium (Beispielsweise ein Homogenitätskriterium) erfüllen, zu verstehen sind, verringern den Anteil an falsch klassifizierten Flächen gegenüber klassischen pixelbasierten Ansätzen und liefern weitaus homogenere Klassifikationsergebnisse [Meinel et al., 2001]. Für die Klassifikation von hochauflösenden Satellitenbildern hat sich daher der segmentbasierte Ansatz gegenüber klassischen pixelbasierten Verfahren bewährt [Neubert, 2005], [Kux und Pinho, 2006], [Blaschke, 2010]. Die Bildsegmente können hier auf Grundlage von Textureigenschaften, spektraler Information, Größe, Form und ihrer Nachbarschaft beschrieben werden und daraufhin zu semantischen Klassen zugewiesen oder zu semantischen Objekten zusammengefasst werden.

Grundlage für eine segmentbasierte Klassifikation ist das Ergebnis einer Bildsegmentierung, also einer Zerlegung eines Bildes in sich nicht überlappende Bildbereiche. Segmentierungsalgorithmen können in die Gruppen merkmalsraumbasierte, bildraumbasierte und hybride Verfahren eingeteilt werden. Merkmalsraumbasierte Verfahren, wie beispielsweise iterative *clustering*-Algorithmen, werden sowohl für unüberwachte Klassifikationen, als auch für die Bildsegmentierung angewendet. Das bekannte Verfahren *iterative self-organizing data analysis technique* (ISODATA) [Bow, 1992] wird häufig für das *clustering* von Farbräumen angewendet. Ein bildraumbasiertes Verfahren ist die schnelle Wasserscheiden-Transformation [De Smet und Pries, 2000]. Hier werden die Grauwerte eines Bildes als Reliefhöhe interpretiert. Wird dieses Relief geflutet, ergeben sich so genannte Wasserscheiden respektive Grenzen zwischen den

gefluteten Becken, welche als Segmentgrenzen interpretiert werden. Eine weitere oft angewendete Methode ist die Multiskalen-Segmentierung nach [Baatz und Schäpe, 2000], welche in der Software eCognition der Firma Trimble [Trimble, 2010] implementiert ist. Hier werden Pixel in Abhängigkeit von zuvor definierten Parametern (bezüglich Segmentgröße, Form und spektraler Information) sukzessive zusammengefasst. Des Weiteren wird durch das Zusammenfassen von sich ähnelnden Bildsegmenten eine Segmenthierarchie generiert. Je nach Bildgröße, Bildinhalt und Wahl der Parameter ist dabei jedoch mit langen Prozessierungszeiten zu rechnen, wobei zusätzlich Expertenwissen bei der Wahl der Parameter, welche die Größe und Form der Segmente beeinflussen, erforderlich sind.

Die Verwendung einfacher quadratischer Bildsegmente wird oft erfolgreich im Bereich Computer Vision [Kumar und Hebert, 2003a], [Awasthi et al., 2007] sowie auch vereinzelt für Fragestellungen in der Fernerkundung [Kolaczyk et al., 2005], [Dai et al., 2009], [Hoberg und Rottensteiner, 2010] verfolgt. Die Bereitstellung quadratischer Segmente ist mutmaßlich der schnellste Weg der Segmentierung und vor allem vor dem Hintergrund der in Abschnitt 1.2 behandelten Rahmenbedingungen sehr attraktiv für das hier behandelte Anwendungsgebiet. Es ist dabei allerdings zu beachten, dass die Segmente hier ohne Einbezug der Bildinformation generiert werden (*top-down*), wodurch die Mischung verschiedener semantischer Klassen in den Segmenten sehr wahrscheinlich wird.

Die Bildsegmentierung kann als Integration von spektral gestützter Kontextinformation interpretiert werden, da hier mitunter die Nachbarschaft von Bildpixeln betrachtet wird und sich ähnelnde Bildpixel zusammengefasst werden. Darüber hinaus existieren noch weitere und mehr abstrahierende Möglichkeiten des Einbezugs von Kontextinformation.

1.3.2 Kontextinformation und graphenbasierte Modelle

Im Zuge einer wahrscheinlichkeitsbasierten Klassifikation werden die thematischen Klassen häufig mittels multivariater Normalverteilungen parametrisiert. Es wird dabei üblicherweise eine bedingte Unabhängigkeit der Merkmale individueller Bildelemente³ angenommen, was zu einer separaten Klassifikation dieser führt [Bishop, 2006]. Vor allem bei spektral schwer trennbaren Klassen können mit dieser Annahme Mehrdeutigkeiten auftreten, welche Fehlklassifikationen sowie „Salz-und-Pfeffer“-Rauschen im Klassifikationsergebnis bewirken. In Abbildung 1.3 wird die Problematik der lokalen Mehrdeutigkeiten anhand von quadratischen Bildsegmenten verdeutlicht. Das obere Segment repräsentiert den Schatten eines Gebäudes und ist sowohl bei visueller Betrachtung der Darstellung auf der linken Seite der Abbildung, als auch bei einer lokalen *maximum likelihood*-Klassifikation (ML) nicht von der thematischen Klasse *Wasser* zu unterscheiden. Der Einbezug von Kontextinformation vermag diese Mehrdeutigkeit, beispielsweise durch das Vorhandensein von Vegetation oder Merkmalen städtischer Bebauung in einer definierten Nachbarschaft des betrachteten Bildelements, aufzulösen.

Verfahren der kontextuellen Klassifikation nutzen Informationen zwischen räumlich benachbarten Pixeln oder Bildelementen, um Klassifikationsergebnisse zu verbessern [Hubert-Moy et al., 2001], [Magnussen et al., 2004]. Graphenbasierte Modelle kombinieren die beiden Gebiete Graphen- und Wahrscheinlichkeitstheorie, wodurch sowohl *a priori* Wissen über die kontextuelle Struktur der thematischen Klassen, als auch statistische Modelle des Bildverhaltens in einem konsistenten mathematischen Rahmen integriert werden können. Die Ableitung der Klassenzugehörigkeiten der Bildelemente wird *Inferenz* genannt und durch die Optimierung

³In dieser Arbeit wird sowohl ein einzelnes Pixel als auch ein Bildsegment als Bildelement bezeichnet.

einer definierten Zielfunktion erreicht.

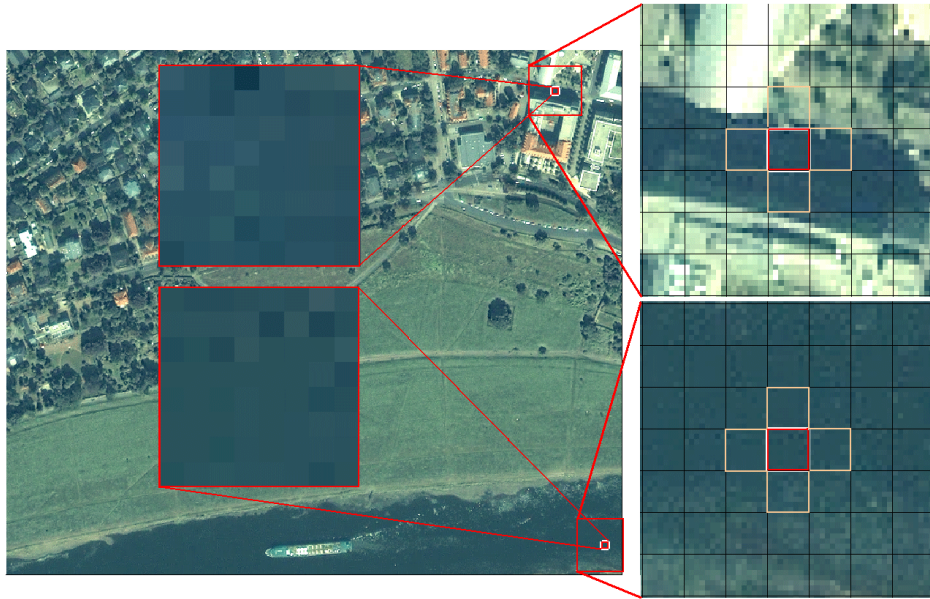


Abbildung 1.3: Zwei Bildelemente sind auf der Grundlage ihrer lokalen spektralen Information nicht voneinander unterscheidbar, obwohl sie eindeutig verschiedenen thematischen Klassen angehören (links). Nachbarschaftsinformation (die gelben Segmente kennzeichnen eine Nachbarschaft erster Ordnung) kann bei der Klassifikation zur Auflösung solcher Mehrdeutigkeiten beitragen (rechts).

Graphentheoretisch betrachtet kann der Einbezug der unmittelbaren lokalen Nachbarschaft der Bildelemente (Nachbarschaft erster oder zweiter Ordnung) im oben geschilderten Beispiel durch ein gitterbasiertes Modell erfolgen. Wie in Abbildung 1.3 rechts zu erkennen ist, weisen fünf der acht unmittelbar benachbarten Bildelemente des oberen Bildelements dieselben bzw. sehr ähnliche spektrale Eigenschaften auf und es wäre möglich, dass noch mehr oder alle Bildelemente der Nachbarschaft diese Eigenschaften haben. Der Einbezug von Kontextinformation mit dieser Nachbarschaftsstruktur würde demnach unter Umständen nicht zu einer Auflösung dieser Mehrdeutigkeit führen. Eine Erweiterung des Modells auf eine größere Nachbarschaft ist prinzipiell möglich, führt allerdings zu einem rapiden Anstieg sowohl der Zahl der Modellparameter als auch der Komplexität und Rechenzeit der Inferenz. Dies verdeutlicht sowohl die Lokalität als auch die Beschränkung des Ansatzes auf einen festen Skalenbereich.

Es wird oft konstatiert, dass ein zu modellierendes Phänomen (z.B. semantische Objekte oder Klassen) ein bestimmtes charakteristisches Verhalten über eine Spanne von Skalen aufweist [Willsky, 2002], [Hay et al., 2003]. Für die Charakterisierung der Landbedeckung auf der Grundlage von Fernerkundungsdaten haben Studien gezeigt, dass die erreichte Genauigkeit der Klassifikation signifikant von der räumlichen Abdeckung der zu Grunde liegenden Bildpixel abhängt [Marceau et al., 1994]. Der Grundgedanke der Multiskalenprozessierung wurde in der Vergangenheit beispielsweise für die Klassifikation anthropogener und natürlicher Objekte [Lang und Blaschke, 2003] sowie für die automatische Detektion von komplexen und zusammengesetzten Objekten [Sun et al., 2010] in hochauflösenden Satellitenbildern aufgegriffen. Wird als Graph an Stelle eines planaren Gitters beispielsweise eine Quadtree-Baumstruktur

für die Definition der Nachbarschaft der Bildelemente verwendet, ergibt sich ein hierarchisches Modell, bei dem jedes innere Bildelement ein Elternelement sowie vier Kindelemente hat. Das obere Bildelement in Abbildung 1.3 ist in dieser Baumstruktur ein Knoten in Level N (vgl. Abbildung 1.4). Die Bildelemente in den größeren Levels (respektive Auflösungen) $N - 1$ bis

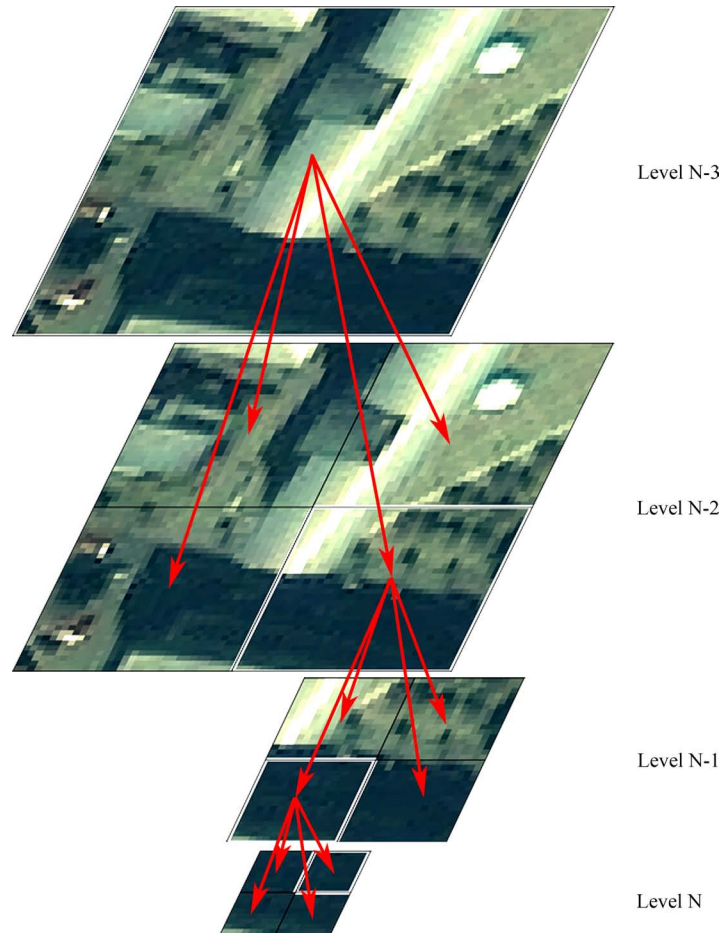


Abbildung 1.4: Definition der Nachbarschaft der Bildelemente mittels Quadtree-Baumstruktur. Dargestellt ist ein Subgraph.

$N - 3$ repräsentieren einen größeren Ausschnitt des Bildes, wobei das betrachtete Bildelement in Level N jeweils mit enthalten ist. In diesen sind sowohl Vegetation als auch Strukturen, welche auf anthropogen beeinflusste Bereiche schließen lassen, enthalten. Verknüpft mit der Bedingung, dass das Auftreten von Wasserflächen in anthropogen beeinflussten Bereichen eine sehr geringe Wahrscheinlichkeit hat, kann diese Information die Fehlklassifikation des Bildelements in Level N vermeiden.

1.4 Zielsetzung der Arbeit

Methoden der Bildanalyse für Anwendungen im Bereich Notfall- und Krisenkartierung sind in der Regel für sehr spezielle Aufgaben und Rahmenbedingungen konzipiert und anwendbar. Die strengen Zeitvorgaben für die Generierung der verschiedenen Produkte sowie die Schwierigkeit der rechnergestützten automatisierten Analyse hochauflösender Satellitenbilder führt dazu, dass Kartierungen auch heutzutage oft auf der Grundlage visueller Interpretationen erfolgen. Daher besteht eine Notwendigkeit der Entwicklung neuer und im Hinblick auf Robustheit, Laufzeit sowie Übertragbarkeit verbesserter Verfahren für die Analyse hochauflösender optischer Satellitenbilder. Auch wenn sich grundsätzlich ein klarer Trend hin zu automatisierten Methoden abzeichnet, wird in dieser Arbeit ein interaktiver Ansatz verfolgt, der als Unterstützungssystem zur Minimierung manueller Interpretationsschritte zu sehen ist. Diese Herangehensweise zielt auf eine Anwendung und Übertragbarkeit der Methode auf verschiedenste Bildprodukte sowie thematische Fragestellungen.

Das Ziel der vorliegenden Dissertation ist die Konzipierung, Evaluierung und prototypische Implementierung eines allgemeinen Rahmenwerks zur (Vor-) Klassifikation hochauflösender optischer Satellitenbilder für Anwendungen im Bereich Notfall- und Krisenkartierungen. Die wahrscheinlichkeitsbasierte Zuordnung quadratischer Bildelemente zu den thematischen Klassen soll durch einen hierarchischen Ansatz der statistischen graphenbasierten Bildmodellierung erfolgen. Es handelt sich dabei um ein hybrides Modell, bei dem ein gerichteter hierarchischer sowie ein ungerichteter planarer Graph miteinander kombiniert wird, so dass sowohl Bildinformation aus verschiedenen Abstraktionsebenen, als auch hierarchische und räumliche Kontextinformation integriert werden kann. Diese Kombination zielt auf den Erhalt der positiven Eigenschaften des hierarchischen Modells hinsichtlich der Laufzeit der Inferenzverfahren. Im Gegensatz zu anderen in der Literatur vorgestellten Ansätzen der graphenbasierten Bildmodellierung wird durch diese sequentielle Vorgehensweise die Prozessierung von Satellitenbildern ohne initial vorhandene Trainingsdaten ermöglicht. Die Schwerpunkte der Arbeit liegen sowohl bei der Konzipierung und Implementierung des Gesamtsystems, als auch speziell bei dem Thema der graphenbasierten Bildmodellierung.

Zunächst soll eine vergleichende Gegenüberstellung existierender Verfahren der kontextuellen Klassifikation vorgenommen werden und daraufhin die Entscheidung zu Gunsten der graphenbasierten Modellierung begründet werden. Anschließend wird ein umfassender Überblick über den Themenbereich der graphenbasierten Bildmodellierung gegeben, wobei auf die prinzipiellen Unterschiede der Verfahren eingegangen werden soll. Insbesondere werden die generellen Unterschiede von planaren und hierarchischen sowie von generativen und diskriminativen Modellen herausgestellt und diskutiert, woraufhin die Anwendung eines generativen Ansatzes in dieser Arbeit begründet werden kann.

Unter Einbezug der hier dargestellten Rahmenbedingungen und Anforderungen soll in dieser Arbeit ein hierarchisches generatives Modell angewendet werden. Diesbezüglich sind alle Komponenten dieses Modells, respektive das Daten- und das *a priori* Modell, zu beschreiben. Des Weiteren werden die statistischen Eigenschaften des Modells sowie die jeweiligen Methoden der Parameterschätzung und Inferenz detailliert dargestellt. Eine nachteilige Eigenschaft des vorgeschlagenen Modells ist mit dem starren Graphen (Quadtree) und den quadratischen Bildsegmenten gegeben. Die Untersuchungen in dieser Arbeit belegen, dass Fehlklassifikationen aufgrund von spektralen Mehrdeutigkeiten (beispielsweise in Bereichen von Klassenübergängen) mit diesem Modell nur begrenzt verhindert werden können. In dieser Arbeit werden daher

Modifikationen der ursprünglichen Inferenzverfahren vorgeschlagen, welche diese Mehrdeutigkeiten mitunter massiv eindämmen können.

Nach der eingehenden Behandlung der Methodik soll diese in ein interaktives Rahmenwerk integriert werden. Die komplette Prozesskette sowie die Möglichkeiten der Bildprozessierung durch Kombination unterschiedlicher Einzelkomponenten soll hier vorgestellt und ein Überblick bezüglich der Anwendung der Methode sowie der dabei zu tätigen interaktiven Eingaben gegeben werden.

Die Bildmerkmale und die Merkmalsreduktion sind entscheidende Aspekte hinsichtlich der Übertragbarkeit und Performanz der Methode. Aus diesem Grund sollen sowohl diverse in der Literatur vorgestellte Merkmale sowie ein Verfahren der Merkmalsselektion in die Methode integriert werden.

Vor dem Hintergrund einer Anwendung des Verfahrens in einem operationellen Kartierungsdienst sollen die einzelnen Prozessschritte hinsichtlich ihrer Sensitivität gegenüber den jeweils einfließenden Parametern empirisch untersucht werden. Insbesondere sollen diesbezüglich die Methoden der Parameterschätzung für das Daten- und das *a priori* Modell behandelt sowie der Mehrwert der Multiskalenprozessierung und der modifizierten Inferenzverfahren mit dem vorliegenden Modell herausgestellt werden.

Schließlich wird das Rahmenwerk durch einen Vergleich mit zwei weiteren Klassifikationsmethoden bezüglich dessen Mehrwert, Effektivität und Übertragbarkeit auf Satellitenbilder mit unterschiedlichen Bildinhalten untersucht.

1.5 Gliederung der Arbeit

Die Arbeit ist wie folgt aufgebaut. In Kapitel 2 werden Verfahren der kontextuellen Klassifikation behandelt und gegenüber gestellt. Diese Übersicht bildet die Grundlage für die Wahl eines Ansatzes der statistischen graphenbasierten Modellierung für diese Arbeit. Anschließend wird ein umfassender Überblick über verschiedene in der Literatur vorgestellte Verfahren der graphenbasierten Bildmodellierung gegeben. Dabei wird vorerst nicht auf die mathematischen, sondern vielmehr auf die anwendungsbezogenen Eigenschaften der Verfahren sowie deren prinzipielle Unterschiede eingegangen und die grundsätzlichen Probleme der existierenden Verfahren bezüglich der Anwendung im Bereich Notfall- und Krisenkartierung herausgestellt.

In Abschnitt 3 werden die Ansätze der graphenbasierten Bildmodellierung eingehend behandelt. Dazu erfolgt eine Einordnung dieser in die Gruppen kausale und nicht kausale Verfahren, wobei jeweils planare und hierarchische Ansätze existieren. Die jeweiligen statistischen Eigenschaften sowie die Methoden der Parameterschätzung und Inferenz werden für jede der vier Gruppen anhand konkreter Beispiele behandelt. Auf der Grundlage der allgemeinen Darstellung der Methoden sowie der mathematischen Eigenschaften dieser kann anschließend die Wahl eines Ansatzes für diese Arbeit erfolgen. Insbesondere erfolgt in diesem Abschnitt eine Diskussion über die theoretischen Unterschiede sowie die praktische Anwendung von diskriminativen und generativen Verfahren der Bildmodellierung im Kontext von Notfall- und Krisenkartierungen.

In Abschnitt 4 wird das hier vorgestellte hybride Modell, dessen statistische Eigenschaften sowie die Methoden der Parameterschätzung und Inferenz behandelt. Es handelt sich dabei um eine Kombination eines hierarchischen gerichteten und eines planaren ungerichteten Modells, mit dem hierarchischer Kontext und Bildinformation aus mehreren Skalenbereichen

sowie räumlicher Kontext zwischen benachbarten Bildelementen innerhalb der feinsten Skala modelliert werden können. Aufgrund von nachteiligen Eigenschaften, welche auf die starre Graphenstruktur des hierarchischen Modells zurückzuführen sind, wird in Abschnitt 4.5.3 eine modifizierte Inferenzmethode vorgestellt. Um die positiven Eigenschaften bezüglich der Rechenzeit der hierarchischen Inferenz zu erhalten, werden die beiden Modelle sequentiell angewendet. In Abschnitt 4.7 wird diesbezüglich diskutiert, inwieweit die beiden Modelle auf gewinnbringende Weise miteinander kombiniert werden können.

Die zuvor behandelte Methodik wird in Abschnitt 5 in ein interaktives Rahmenwerk integriert. Hier wird der gesamte Ablauf der Prozessierung vorgestellt und exemplarisch erläutert. Insbesondere werden hier die verschiedenen Möglichkeiten der Bildprozessierung, die jeweils vom Bildinterpreten zu tätigenen Eingaben und Definitionen von Parametern sowie die Möglichkeiten der Visualisierung von Zwischenergebnissen dargestellt.

In Abschnitt 6 erfolgt die konkrete Anwendung und empirische Untersuchung des Rahmenwerks am Beispiel von sechs verschiedenen Subszenen des Sensors IKONOS, welche zunächst in Abschnitt 6.1 vorgestellt werden. Die verwendeten Bildmerkmale werden in Abschnitt 6.2 vorgestellt und die Identifikation der problemrelevanten Bildmerkmale wird in Abschnitt 6.3 behandelt. Anschließend werden die Eigenschaften des Rahmenwerks mittels verschiedener Experimente untersucht. In den Abschnitten 6.4 und 6.5 wird die Sensitivität der Parameterschätzung für das Daten- und *a priori* Modell bezüglich der jeweils einfließenden Parameter und Trainingsdaten untersucht. Die semiautomatische Bereitstellung der Trainingsdaten für das *a priori* Modell wird in Abschnitt 6.6 behandelt. In den Abschnitten 6.7 und 6.8 wird der Mehrwert der Multiskalenprozessierung mit dem vorliegenden Modell herausgestellt und die Ergebnisse der Methode mit denen einer ML-Klassifikation sowie einer weiteren hierarchischen Inferenzmethode (*sequential maximum a posteriori* Inferenz (SMAP-Inferenz)) verglichen. Anhand dieses Vergleichs wird die Relevanz des Rahmenwerks und vor allem der Mehrwert der in Abschnitt 4 behandelten modifizierten Inferenzmethoden deutlich herausgestellt. Abschließend wird in Abschnitt 6.9 der Zusammenhang zwischen einer Konfidenzkarte der Klassifikation und den tatsächlichen Fehlklassifikationen herausgestellt und damit die Aussagekraft und Verlässlichkeit dieser Information untersucht.

Eine Zusammenfassung und Bewertung der Ergebnisse dieser Arbeit sowie ein Ausblick bezüglich weiterer Forschungsarbeiten in dem vorliegenden Themengebiet wird in Abschnitt 7 gegeben.

2 Kontextbasierte Klassifikation

Bei einer ML-Klassifikation erfolgt eine Parametrisierung der thematischen Klassen häufig durch multivariate Normalverteilungen in einem Merkmalsraum, wobei die Parameter dieser aus Trainingsdaten geschätzt werden. Auf der Grundlage der jeweils beobachteten spektralen Werte sowie der Parametrisierung wird jedes Bildpixel durch ein daraus berechnetes statistisches Ähnlichkeitsmaß klassifiziert. Die Klassifikationsgüte dieser pixelbasierten Methode ist stark abhängig von weißem Rauschen, welches die Ähnlichkeit thematischer Klassen im Merkmalsraum begünstigen kann [Sharma und Sarkar, 1998], [Kartikeyan et al., 1994]. Kontextinformation kann solche Mehrdeutigkeiten auflösen und somit bei der Unterscheidung zwischen thematischen Klassen mit ähnlichen spektralen Eigenschaften sehr hilfreich sein.

Der Begriff Kontext bezieht sich prinzipiell nicht ausschließlich auf eine räumliche Nachbarschaft von Bildelementen. Vielmehr entscheidet die üblicherweise durch einen Graphen definierte Nachbarschaft über die Art des Kontexts. In Abschnitt 1.3.2 wird ein hierarchischer Baum als Graph verwendet (Abbildung 1.4). Das daraus resultierende lokale Nachbarschaftssystem eines Bildelements bezieht sich dabei nicht auf die direkten räumlichen Nachbarn, sondern auf die Nachbarn in den angrenzenden Hierarchieebenen. Daher kann hier von hierarchischem Kontext gesprochen werden. Die kontextuellen Relationen, die in dieser Arbeit behandelt werden, sind ausschließlich räumlicher oder hierarchischer Natur. Darüber hinaus existieren weitere kontextuelle Relationen, wie temporaler Kontext (für Zeitreihenanalysen oder Änderungsdetektionen) oder auch Kontext bezüglich unterschiedlicher Daten zu einem gleichen Aufnahmezeitpunkt und mit gleicher räumlichen Abdeckung (Datenfusion beispielsweise von optischen und SAR-Daten). In diesem Abschnitt wird ein Literaturüberblick über existierende Verfahren der kontextbasierten Klassifikation sowie speziell über Verfahren der graphenbasierten Bildmodellierung gegeben.

2.1 Verfahren zur kontextbasierten Klassifikation

Verfahren der kontextuellen Klassifikation verwenden Glättungstechniken, räumliche Statistiken, *fuzzy-logic*, Segmentierungen, neuronale Netze oder Markov-Zufallsfelder [Lu und Weng, 2007].

Filter werden sowohl in Vor- als auch in Nachbearbeitungsschritten auf das Originalbild respektive auf das Ergebnis einer Klassifikation angewendet. Ein Beispiel für letztere ist der *majority*-Filter [Mather, 2004], der eine Glättung des ursprünglichen Ergebnisses bewirkt. Ein Standard *majority*-Filter weist einem Pixel die Klasse zu, welche in seiner Nachbarschaft am häufigsten vorkommt. Eine weitere Möglichkeit der Nachprozessierung ist mit der Anwendung morphologischer Operatoren, wie *erosion*, *dilation*, *opening* oder *closing* gegeben.

Die Erfassung und Beschreibung der lokalen Variationen in den spektralen Bändern eines Satellitenbildes kann ebenfalls durch Filterung erfolgen. Hier werden beispielsweise Texturmerkmale abgeleitet, die als zusätzliche spektrale Bänder in die Klassifikation einfließen. Dafür werden geostatistische Techniken [Atkinson und Lewis, 2000] sowie räumliche autokorrelationsstatistische Informationen [Myint et al., 2007] herangezogen.

Die Filterung oder Glättung von Klassifikationsergebnissen kann zu einer Verbesserung der Klassifikationsgüte führen. Bevor jedoch die Anwendung solcher Nachbearbeitungsschritte in dieser Arbeit erwogen wird, soll zunächst das Klassifikationsverfahren selbst behandelt werden, da dadurch möglicherweise Nachbearbeitungsschritte dieser Art obsolet werden. Die Anwendung von Texturmerkmalen kann hingegen besonders bei hochauflösenden Bildern und in strukturierten Bildbereichen, die beispielsweise anthropogen beeinflusste Gebiete repräsentieren, hilfreich bei der Unterscheidung thematischer Klassen sein. Daher sollen hier ebenfalls Texturmerkmale verschiedenster Art für die Klassifikation bereitgestellt werden.

An Stelle einer harten Zuweisung einer Klasse zu einem Pixel kann mittels *fuzzy* Klassifikationsmethoden eine „weiche“ Entscheidung getroffen werden. Unter Verwendung einer so genannten *fuzzy partition* Matrix kann nach [Wang, 1990] der Grad der Zugehörigkeit eines Pixels zu jeder thematischen Klasse repräsentiert werden. Dieser Grundgedanke wurde in [Cai et al., 2007] mit einer unüberwachten Segmentierung durch *clustering* des Merkmalsraums kombiniert. Das Ergebnis ist eine Familie schneller und robuster *fuzzy c-means clustering* Algorithmen, bei denen die Zugehörigkeit eines Pixels zu mehreren Clustern möglich ist. Neben spektralen Informationen werden hier auch räumliche Informationen bezüglich der Nachbarschaft eines Pixels im Bildraum einbezogen, wodurch Detailerhalt sowie Robustheit gegenüber Rauschen erreicht wird. Ein weiterer Ansatz, bei dem die *fuzzy-set* Theorie Anwendung findet, ist mit der Software *eCognition* gegeben. Hier wird eine segmentbasierte Multiskalenprozessierung ermöglicht, bei der kontextuelle Bedingungen mit einer hierarchischen semantischen Wissensbasis definiert werden. Die Klassifikation von Landbedeckung oder gar Landnutzung erfolgt hier auf der Grundlage vielfältiger spektraler, textueller, formbasierter und kontextueller Merkmale, wobei jeweils *fuzzy* Funktionen und logische Regeln bezüglich einzelner Merkmale für die Trennung zwischen den thematischen Klassen herangezogen werden. Ein entscheidender Nachteil bei diesem Ansatz besteht in der Schwierigkeit der Erstellung eines solchen Regelwerks, da die Merkmale, *fuzzy* Funktionen, Schwellwerte und nicht zuletzt die Reihenfolge der Anwendung dieser manuell gewählt werden müssen. Die Übertragbarkeit eines definierten Regelsatzes auf unterschiedliche Daten und Problemstellungen ist meist mit einem enormen Aufwand bezüglich der Anpassung der Regelsätze verbunden.

Wie in Abschnitt 1.3.1 erläutert wurde, kann die Bildsegmentierung ebenfalls als Ansatz für den Einbezug von Kontext gesehen werden. Ziel dieser *low-level* Operation ist die Unterteilung des Bildes in sich nicht überlappende Segmente. Es kann zwischen merkmalsraumbasierten Verfahren und bildraumbasierten Verfahren unterschieden werden. Erstere sind beispielsweise *clustering*-Verfahren, wie die ISODATA Methode sowie auf Histogrammschwellwerten basierende Techniken. Im Bildraum operierende Verfahren teilen und verschmelzen (*split and merge*) Bildelemente oder lassen diese anwachsen (*region growing*), wodurch eine räumliche Kompaktheit der Segmente gewährleistet wird. Wie zuvor beschrieben, wird der segmentbasierte Ansatz erfolgreich für die Klassifikation von Landbedeckung in hochauflösenden Satellitenbildern angewendet und daher auch in dieser Arbeit verfolgt.

Neuronale Netze fallen ebenfalls in die Gruppe der im Bildraum arbeitenden Verfahren. Grundsätzlich handelt es sich hier um Strukturen, welche aus einer großen Anzahl stark vernetzter elementarer Prozessoren bestehen. Neben dem Vorteil der parallelen Prozessierung ist hier auch der Einbezug von räumlichem Kontext möglich. Es handelt sich um einen nicht parametrischen Ansatz, der ohne Annahmen über die Verteilung der Bilddaten im Merkmals-

raum aus kommt. Weitere Vorteile eines solchen Klassifikators sind die gute Lernfähigkeit sowie die Fähigkeit zur Generalisierung. Nachteilig ist das langsame, aufwändige und iterative Training aufgrund der hohen Konnektivität der Elementarprozessoren sowie die Forderung nach umfassenden Trainingsdaten. Zudem existieren keine expliziten Regeln für die Wahl einer adäquaten Netzarchitektur, wodurch sich dieser Ansatz nicht für diese Arbeit eignet. Weiterführende Informationen zu diesem Thema sind beispielsweise in [Lucchese und Mitra, 2001] und [Pal und Pal, 1993] zu finden.

Markov-Zufallsfelder (MRF) gehören der Klasse der graphenbasierten Ansätze an und werden für die kontextuelle Klassifikation am häufigsten angewendet [Magnussen et al., 2004]. Es handelt sich dabei um Bayes'sche statistische Modelle, bei denen die natürliche Integration von sowohl statistischen Modellen des Bildverhaltens als auch *a priori* Wissen über die kontextuelle Struktur von semantischen Klassen erfolgen kann. Ein Graph definiert hier auf einfache und übersichtliche Art die Abhängigkeiten zwischen den Bildelementen. Um rechnerische Tragbarkeit zu gewährleisten, werden neben dieser Lokalität üblicherweise vereinfachende Annahmen getroffen, so dass eine Zerlegung globaler Wahrscheinlichkeitsverteilungen in eine (teil-) faktorisierte Form und damit die Entwicklung effizienter Algorithmen für die Berechnung der gesuchten Größen ermöglicht wird. Verschiedenste Aufgabenstellungen, wie beispielsweise Bildrekonstruktion, Kantendetektion, Klassifikation und Verminderung von Rauschen können mit graphenbasierten Ansätzen bearbeitet werden. Aufgrund von Unsicherheiten, wie Rauschen und spektralen Mehrdeutigkeiten (vergleiche Abschnitt 1.3), existiert keine exakte oder perfekte Lösung eines Problems. Mit statistischen graphenbasierten Ansätzen kann die Bestimmung einer, im Sinne der stochastischen Modellbildung, optimalen jedoch inexakten Lösung erfolgen [Li, 2009]. Die Algorithmen dafür basieren auf fundierten mathematischen Prinzipien und nicht auf ad-hoc Heuristiken [Chellappa und Jain, 1993]. Im Vergleich zu anderen Verfahren, wie der auf *fuzzy-sets* basierten Klassifikation, können hier die jeweiligen Modellparameter auf der Grundlage von Trainingsdaten durch entsprechende Algorithmen für jedes Klassifikationsproblem gesondert geschätzt werden, so dass eine manuelle Anpassung dieser nach dem *trial-and-error* Prinzip entfällt. Aus diesem Grund wird in dieser Arbeit ein statistischer graphenbasierter Ansatz gewählt. Im Folgenden wird eine Literaturübersicht zu diesem Thema gegeben.

2.2 Graphenbasierte Bildmodellierung

In den letzten vier Jahrzehnten wurden in den Bereichen *computer vision* (maschinelles Sehen) und Fernerkundung viele verschiedene Ansätze zur kontextbasierten Klassifikation mittels graphenbasierter Bildmodellierung vorgeschlagen. Prinzipiell handelt es sich bei diesen Modellen um globale Wahrscheinlichkeitsverteilungen, welche bezüglich eines Graphen definiert sind. Der Graph definiert die lokale Nachbarschaft der Bildelemente, wodurch die Abhängigkeiten dieser sehr leicht ablesbar ist. Da somit im Umkehrschluss auch die essentielle Eigenschaft der Unabhängigkeit definiert ist, wird auch von Unabhängigkeitsgraphen gesprochen.

Die Modelle werden in der Literatur anhand ihrer Eigenschaften jeweils in zwei Gruppen eingeteilt. So kann eine Untergliederung in die Gruppen diskriminativ und generativ, zyklisch und azyklisch, gerichtet und ungerichtet oder hierarchisch und planar erfolgen. Da an dieser Stelle lediglich eine übergeordnete und anwendungsbezogene Betrachtung existierender Ansätze erfolgen soll, ohne detailliert auf die mathematischen Eigenschaften und Unterschiede

dieser einzugehen, wird in diesem Abschnitt eine Gruppierung nach hierarchischen und planaren Modellen vorgenommen. Auch vor dem Hintergrund einer Multiskalenprozessierung in dieser Arbeit erscheint diese Einteilung als sehr dienlich.

2.2.1 Planare Modelle

In planaren Modellen werden die Abhängigkeiten zwischen den Bildelementen durch einen flachen Graphen definiert. Das gitterbasierte MRF ist das am häufigsten in der Literatur angewendete Modell. Es wurde im Bereich *computer vision* durch die Arbeiten von Cross und Jain [Cross und Jain, 1983], Besag [Besag, 1974] [Besag, 1986] sowie Geman und Geman [Geman und Geman, 1984] bekannt. MRFs werden in einem wahrscheinlichkeitsbasierten generativen Rahmenwerk verwendet, welches die gemeinsame Verteilung der beobachteten Variablen (Bildinformation) und der dazu korrespondierenden Klassenzugehörigkeiten modelliert [Li, 2009]. Erfolgt bei einem binären Klassifikationsproblem die Definition der Nachbarschaft durch ein Gitter, welches jedes Bildelement beispielsweise mit seinen vier unmittelbaren Nachbarn (Nachbarschaft erster Ordnung) verbindet (vergleiche Abbildung 2.1), sowie eine Verwendung richtungs- und ortsunabhängiger sowie paarweiser Potentiale zwischen den Elementen, spricht man von einem Ising-Modell. Die lokale Beschränkung der Nachbarschaft wird dadurch legi-

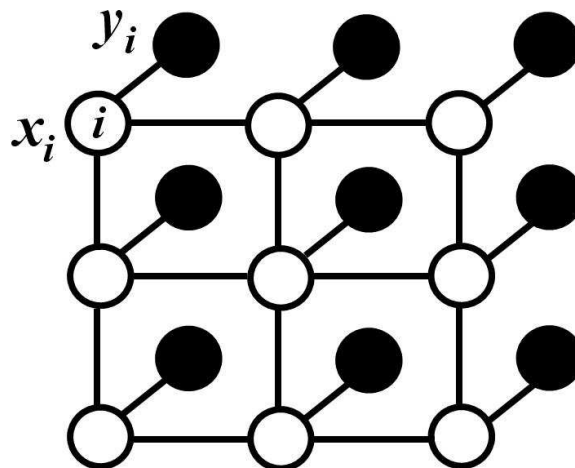


Abbildung 2.1: Schematische Darstellung eines Markov-Zufallsfeldes aus [Kumar, 2005]. Jedes Pixel i wird mit zwei Zufallsvariablen assoziiert. Die dunklen Knoten repräsentieren die Beobachtungen y_i , die hellen Knoten die gesuchten Klassenzugehörigkeiten x_i .

timiert, dass mit einer wachsenden Entfernung die Korrelation zwischen den Bildelementen rapide abnimmt. Die Favorisierung von Klassengleichheit zwischen einem Element und dessen Nachbarn mit dem Ising-Modell (Zweiklassenproblem) oder dem Potts-Modell [Bouman und Shapiro, 1994] bewirkt eine Glättung. Ein üblicherweise manuell definiertes Gewicht reguliert dabei den Einfluss des Kontexts.

In [Bruzzone und Prieto, 2000] wird ein MRF auf Differenzbilder von Landsat 5-Szenen (räumliche Auflösung: 30 m) angewendet, um eine unüberwachte kontextbasierte Veränderungsdetektion durchzuführen. Der Einbezug von temporalem Kontext kann neben der Prozessierung von Differenzbildern auch durch eine entsprechende Erweiterung des Modells

erfolgen. Daraus resultieren für zwei Szenen, welche den gleichen Bereich der Erdoberfläche zu unterschiedlichen Zeitpunkten repräsentieren, zwei planare MRFs, wobei jeweils zusätzlich die beiden korrespondierenden Knoten verbunden sind. In [Liu et al., 2008] wird durch eine solche Modellerweiterung eine Veränderungsdetektion in Waldgebieten unter Verwendung von Landsat Daten durchgeführt.

Die Knoten in einem MRF sind durch einen ungerichteten Graphen miteinander verbunden. *Hidden Markov-Modelle* (HMMs) basieren im Gegensatz dazu auf Markov-Ketten [Markov, 1913]. An Stelle von paarweisen Cliquenpotentialen¹ werden Übergangswahrscheinlichkeiten, welche von den Zuständen der benachbarten Elemente in der Kette abhängen, herangezogen. In [Li et al., 2000] wird dieser Ansatz auf zwei Dimensionen erweitert. Dabei wird das gesamte Bild sukzessive diagonal durchquert und jeweils die schon besuchten Nachbarn der quadratischen Bildsegmente einbezogen. Die Relevanz des Ansatzes wurde anhand der Segmentierung von Dokument- sowie Luftbildern demonstriert. Markov-Ketten in Form von *Hilbert-Peano scans* werden in [Derrode und Pieczynski, 2000] ebenfalls für die Segmentierung von Dokumenten und Luftbildern angewendet. Auch hier kann eine Erweiterung des Modells auf räumlich-temporalen Kontext erfolgen. Grundsätzlich ist zu beachten, dass ein eindimensionaler Ansatz, bei dem eine bestimmte Anordnung der Elemente angenommen wird, üblicherweise suboptimal ist [Li et al., 2000] und das Ergebnis generell von der Wahl des Graphen abhängt. Dieser Sachverhalt ist für alle graphenbasierten Modelle gleichsam gültig.

In [Zhong et al., 2007] wird ein MRF-basiertes Rahmenwerk mit adaptiven Nachbarschaftssystemen vorgeschlagen. Unter Verwendung verschiedener definierter Nachbarschaftssysteme wird nicht die oben beschriebene Glättung erzielt, sondern vielmehr der Erhalt von gewünschten Strukturen, wie beispielsweise Linien, erwirkt. Zusätzlich dazu werden neben dem herkömmlich verwendeten statistischen Datenmodell der thematischen Klassen im Merkmalsraum auch die Variationen zwischen den thematischen Klassen basierend auf den Differenzen zwischen den Merkmalsvektoren modelliert. Das Verfahren wurde erfolgreich für die Klassifikation von Nahbereichsfotografien sowie von Landsat TM-Satellitenbildern angewendet. Bei den bisher genannten generativen Modellen wird die gemeinsame Verteilung $P(x, y)$ und somit auch implizit die beobachteten Bilddaten durch $p(y|x)$ modelliert. Die rechnerische Tragbarkeit wird durch die Annahme der Unabhängigkeit der beobachteten Bilddaten gegeben den Nachbarelementen erwirkt. In [Zhong et al., 2007] wird durch die Formulierung eines Datenmodells, welches auf Statistiken zwischen benachbarten Bildelementen basiert, der zusätzliche Einbezug von Kontextinformation bezüglich der Bilddaten realisiert.

Conditional random fields (CRF) kommen ohne die oben genannten Unabhängigkeitsannahmen aus und wurden erstmals von [Lafferty et al., 2001] im Kontext der Segmentierung und Klassifizierung sequentieller Daten vorgeschlagen. Dieser Ansatz ist im Bereich der Mustererkennung aufgrund der Fähigkeit, Abhängigkeiten nicht nur zwischen den Klassenlabels² benachbarter Bildelemente, sondern auch zwischen den Labels und den Bildmerkmalen zu modellieren, bekannt geworden. In [Kumar und Hebert, 2003a] erfolgt mit den diskriminativen Zufallsfeldern (DRF) erstmals eine Übertragung dieses Ansatzes in den Bereich der Bildverarbeitung, wodurch grundsätzlich die Anwendung von überlappenden Bildmerkmalen ermöglicht wird. Des Weiteren kann, wie oben beschrieben, die Modellierung der Kontextinformation

¹Eine Clique ist ein Subgraph des Bildmodells. In dieser Arbeit werden ausschließlich die sehr weit verbreiteten paarweisen Cliquen verwendet, da diese eine einfache Form aufweisen und daraus ein niedriger Prozessierungsaufwand resultiert.

²Ein Klassenlabel entspricht der diskreten Variable x , die die Klassenzugehörigkeit eines Bildelements beschreibt.

ebenfalls in Abhängigkeit der Bildinformation erfolgen. In [Kumar, 2005] werden mittels DRFs unter anderem künstliche Strukturen in Nahbereichsbildern klassifiziert, wobei auch hier quadratische Segmente betrachtet werden. Um die Charakteristika dieser Strukturen zu erfassen, wird für jedes zu klassifizierende Bildelement sowie aus größeren Bildbereichen, die das Bildelement zentral beinhalten, ein gradientenbasiertes Maß für dessen Strukturiertheit abgeleitet. In [Hoberg und Rottensteiner, 2010] werden für DRFs ebenfalls gradientenbasierte Merkmale aus verschiedenen Skalen sowie Farbmerkmale für die Klassifikation von Siedlungsbereichen basierend auf IKONOS-Bildern mit einer räumlichen Auflösung von 4 m angewendet. Dabei zeigt sich im Vergleich zu einer lokalen ML-Klassifikation der quadratischen Bildsegmente eine deutliche Verbesserung der Klassifikationsgüte.

Auch wenn in DRFs überlappende Bildbereiche erlaubt sind, beschränkt sich die Interaktion der Bildelemente bei planaren Graphen weiterhin auf eine feste Skala, wie beispielsweise der Pixelebene. Vor allem bei der Prozessierung hochauflösender Daten ist dadurch die Modellierung kontextueller Bedingungen räumlich sehr begrenzt. Die Nachbarschaftssysteme können prinzipiell auf größere Bereiche ausgedehnt werden, allerdings bewirkt dies einen rapiden Anstieg der Modellparameter, sowie erheblich längere Rechenzeiten. Hierarchische graphenbasierte Modelle können die intrinsische hierarchische Natur von Fernerkundungsdaten besser erfassen und weitreichende kontextuelle Beziehungen modellieren [Bouman und Shapiro, 1994].

2.2.2 Hierarchische Modelle

Bei hierarchischen Ansätzen werden häufig kausale azyklische Baumstrukturen als Unabhängigkeitsgraph herangezogen, da sich dadurch generell rechnerisch effiziente, nicht iterative und exakte Inferenzmethoden ableiten lassen [Pérez et al., 2000]. Gerichtete azyklische Graphen sind unter dem Namen Bayes'sche Netze bekannt. Die Bezeichnung dieser Modelle als MRF wurde erstmals in [Bouman und Shapiro, 1994] eingeführt und auch in dieser Arbeit verwendet.

Durch gerichtete Quadrees definierte MRFs werden beispielsweise in [Laferté et al., 2000] und [Collet und Murtagh, 2004] für die Klassifikation von Luftbildern und verrauschten künstlichen Bildern sowie für die Segmentierung von astronomischen Aufnahmen angewendet. Die thematischen Klassen werden bei diesem generativen Ansatz mittels Gaußverteilungen auf verschiedenen Hierarchieebenen respektive Levels des Quadrees modelliert. Aufgrund des gerichteten Graphen werden die kontextuellen Bedingungen an Stelle von Potentialen mittels bedingten Wahrscheinlichkeiten, den so genannten Übergangswahrscheinlichkeiten, zwischen den thematischen Klassen in benachbarten Hierarchieebenen formuliert, so dass hier von hierarchischem Kontext gesprochen werden kann. Bei der Quadtree-Struktur können räumlich benachbarte Elemente innerhalb eines Skalenbereichs von unterschiedlichen Elternelementen in größeren Skalenbereichen abstammen, was zu Blockeffekten in den Inferenzergebnissen führen kann. Um diesen Effekt zu reduzieren, wird in [Bouman und Shapiro, 1994] ein Multiskalen Zufallsfeld (MSRF) zur Bayes'schen Bildsegmentierung basierend auf Pyramidengraphen vorgeschlagen. Durch die erhöhte Konnektivität der Knoten des Baumes ergeben sich wiederum approximative Inferenzmethoden. Weitere Möglichkeiten der Verringerung des Effekts bestehen in einer nachträglichen Glättung des hierarchischen Inferenzergebnisses beispielsweise durch die Anwendung eines planaren MRF Modells in der feinsten Ebene des Quadrees [Luettgen et al., 1993], [Martinis et al., 2011], sowie durch die Verwendung von Baumstrukturen mit überlappenden Bildelementen [Irving et al., 1997]. In [Wilson und Li, 2002] wird hingegen ein Quadtree verwendet, in dem jedes Bildelement nicht nur mit dem Eltern- und den Kindknoten der benachbarten Hierarchieebenen, sondern auch mit den unmittelbaren Nachbarn innerhalb

der jeweiligen Ebene verbunden ist (räumlich-hierarchischer Kontext). Daraus resultiert wiederum ein zyklischer Graph und somit aufwändige und iterative Inferenzmethoden.

Kennzeichnend für nahezu alle in der Literatur vorgestellten hierarchischen Ansätze ist, dass in jedem Skalenbereich dieselben thematischen Klassen modelliert werden. Bezüglich einer Klassifikation von Fernerkundungsdaten erscheint dieser Modellierungsansatz nicht zielführend, da in größeren Skalenbereichen nicht dieselben semantischen Klassen wie beispielsweise in der feinsten Auflösungsstufe repräsentiert werden können. In [Kolaczyk et al., 2005] werden für Zwecke der Landbedeckungsklassifikation so genannte *mixlets* vorgestellt, wodurch beispielsweise das nebeneinander und zyklische Auftreten von Bäumen und Sandboden in höheren Hierarchieebenen zu der Klasse *Baumplantage* aggregiert werden kann. Es wird eine dyadische Baumstruktur für die Repräsentation der Struktur der thematischen Klassen in verschiedenen Auflösungsstufen verwendet. Parallel dazu wird für die Definition der Nachbarschaften zwischen den einzelnen Skalenbereichen wiederum ein Quadtree herangezogen.

Üblicherweise basiert die Modellierung auf regulären Unabhängigkeitsgraphen mit einfachen quadratischen Bildelementen. Für die Generierung dieser Struktur werden die Bilddaten nicht berücksichtigt, so dass das Auftreten von Mischklassen in den Bildelementen wahrscheinlich ist. Um dies zu verhindern, wird in [He et al., 2006] ein planares Modell mit einer so genannten *super pixel* Repräsentation des ursprünglichen Bildes verwendet, was zu einer gewollten Übersegmentierung des Bildes führt. In [Martinis et al., 2011] hingegen werden die Methoden der hierarchischen Modellierung mittels Quadtree auf irreguläre hierarchische Graphen übertragen. Hier wird der Graph durch das Ergebnis einer Multiskalensegmentierung nach [Baatz und Schäpe, 2000] definiert. Die Methode wird erfolgreich und operationell für die Veränderungsdetektion von Hochwasserereignissen aus hoch- und mittelauflösenden SAR-Daten vom Kartierungsdienst ZKI angewendet.

Äquivalent zu den planaren Modellen wurden in der Vergangenheit vereinzelt hierarchische generative Ansätze vorgestellt, um die Unabhängigkeitsannahmen bezüglich der Bilddaten zu lockern. In [Cheng und Bouman, 2001] wird ein Ansatz für die Segmentierung von Dokumenten vorgeschlagen, bei dem für die Merkmale eines Bildelements eine Abhängigkeit sowohl von dessen Klassenzugehörigkeit, als auch von den Bildmerkmalen des Elternsegments angenommen wird. Eine Vorhersage der Merkmale in einem Bildelement wird daraufhin für jede Klasse durch eine lineare Funktion, welche von den Merkmalen des Elternsegments sowie von zwei Koeffizienten abhängig ist, getroffen. Für das Datenmodell wird an Stelle der üblicherweise verwendeten *likelihood* die Differenz aus Merkmalsvektor und Vorhersage verwendet. Die Modellierung der Übergangswahrscheinlichkeiten erfolgt auf der Grundlage eines binären Klassenwahrscheinlichkeitsbaumes. In [Wilson und Li, 2002] werden wiederum für die Segmentierung von texturierten Bildern - ähnlich wie in [Zhong et al., 2007] - auf der Grundlage von Differenzen der Merkmalsvektoren benachbarter Bildelemente die Variationen zwischen den Klassen modelliert. Dafür sind dieselben Bildmerkmale in allen Elementen erforderlich.

Auch im Bereich der hierarchischen Modelle wurden in den vergangenen Jahren diskriminative Verfahren vorgestellt. Bei *multiscale conditional random fields* (mCRF) [Richard et al., 2004], die für die Klassifikation von Nahbereichsfotografien angewendet wurden, wird keine feste hierarchische Struktur, sondern lokale, regionale und globale Klassifikatoren (Neuronale Netze), die miteinander verknüpft sind, verwendet. *Tree structured conditional random fields* [Awasthi et al., 2007] (TCRF) verwenden wiederum einen ungerichteten Quadtree, wodurch sich auch hier nichtiterative Inferenzmethoden ableiten lassen. Die Parameter dieses Modells sind lokale und globale Potentiale, welche den Einfluss der Bilddaten auf die Klassenlabels sowie die Kompatibilität der versteckten Variablen in einem Level des Baumes gegeben

den Nachbarn repräsentieren. Ähnlich wie bei den mCRFs fließen hier die Bilddaten lediglich im feinsten Level des Quadtrees ein. Nach Wissen des Autors wurden diskriminative hierarchische Modelle bisher nicht für die Klassifikation von Fernerkundungsbildern angewendet.

2.3 Diskussion

Die graphenbasierte Bildmodellierung wird für eine Vielzahl unterschiedlicher Anwendungen eingesetzt. Kennzeichnend für nahezu alle Verfahren ist, dass reguläre Graphen sowie rechteckige oder quadratische Bildelemente verwendet werden. Planare pixelbasierte Modelle mit gitterbasierten Graphen erscheinen aufgrund der sehr begrenzten Nachbarschaft vor allem für die Prozessierung hochauflösender Satellitenbilder ungeeignet. Eine Anwendung auf Bildsegmente vermindert diesen Umstand und ermöglicht darüber hinaus die Berechnung von Textur- und anderen Strukturmerkmalen.

Im Gegensatz zu planaren Modellen ermöglichen hierarchische Ansätze den Einbezug von weitreichenderen kontextuellen Interaktionen. Des Weiteren besteht prinzipiell die Möglichkeit des Einbezugs von Bildinformationen unterschiedlicher Skalenbereiche. Vor dem Hintergrund eines möglicherweise skalenspezifischen Verhaltens von Naturerscheinungen in einem Bild sowie der inhärenten hierarchischen Struktur von Satellitenbildern, soll auch in dieser Arbeit ein hierarchischer Ansatz verfolgt werden. Für die hierarchische Bildrepräsentation wird häufig ein gerichteter Quadtree verwendet, wodurch eine Anwendung schneller, nichtiterativer und mathematisch exakter Inferenzmethoden möglich ist. Daher erscheint diese Struktur auch für eine Anwendung im Bereich Notfall- und Krisenkartierungen sehr attraktiv. Die sehr schnelle Generierung eines Quadtrees ist ein weiteres Argument für die Verwendung eines Graphen dieser Form.

In nahezu allen in der Literatur vorgestellten hierarchischen Verfahren (ausgenommen [Kolaczyk et al., 2005]) werden in allen Hierarchieebenen dieselben thematischen Klassen modelliert sowie dieselben Merkmale zur Beschreibung der Bildeigenschaften herangezogen. Diese Vorgehensweise erscheint für die Erfassung und Beschreibung der Bildeigenschaften eines Satellitenbildes nicht sinnvoll. Es ist vielmehr zu erwarten, dass in verschiedenen Skalenbereichen sowie in damit einhergehend unterschiedlich großen Bildelementen jeweils auch andere Bildmerkmale für die Trennung zwischen den Klassen relevant werden.

Diskriminative Ansätze ermöglichen auch bei planaren Modellen die Verwendung von Bildmerkmalen aus mehreren Skalen. Allerdings beschränkt sich die Kontextmodellierung hier weiterhin auf den flachen und üblicherweise gitterförmigen Graphen. Hierarchische diskriminative Modelle, wie die TCRFs [Awasthi et al., 2007], verwenden aus Effizienzgründen wiederum Quadtrees. Die Betrachtung der Bildinformation erfolgt hier jedoch analog zu den DRFs in der feinsten Ebene des Quadtrees in Form von Multiskalen-Merkmalsvektoren. Dies impliziert wiederum, dass die Knoten in den darüber liegenden Skalenbereichen zwar Klassenlabels x , jedoch keine Bildinformationen y repräsentieren.

Im Gegensatz zu den oben angeführten Ansätzen ist die Klassifikation von Satellitenbildern im Kontext von Notfall- und Krisenkartierungen mit sehr restriktiven und einschränkenden Rahmenbedingungen verbunden. Während Verfahren häufig im Kontext einer bestimmte Anwendung vorgeschlagen werden, ist in dieser Arbeit ein hohes Maß an Flexibilität gefordert. Weitere Unterschiede bestehen in der Forderung nach einer schnellen Prozessierungszeit sowie in der sehr eingeschränkten Verfügbarkeit von Trainingsdaten. So wird in dieser Arbeit eine

sehr schnelle Prozessierung auf der Basis von möglichst wenig interaktiv bereitzustellenden Trainingsdaten und anderen Informationen angestrebt, wohingegen in den meisten anderen Anwendungen keine Bedingungen bezüglich der Laufzeit bestehen und ausreichend Trainingsdaten in Form von mehreren klassifizierten Bildern vorhanden sind.

Vor dem Hintergrund der begrenzten Datengrundlage für die Aufgabe der Klassifikation in dieser Arbeit erscheint ein Ansatz der Multiskalenprozessierung sehr vorteilhaft. Bei einer Verwendung eines Quadrees für die Repräsentation des Unabhängigkeitsgraphen ergeben sich effiziente nicht iterative Algorithmen der Inferenz, wodurch begrenzte und verhältnismäßig kurze Rechenzeiten zu erwarten sind. Um die hierarchische Struktur von Satellitenbildern adäquat erfassen zu können, wird im Unterschied zu den existierenden Verfahren die freie Definition von unterschiedlichen semantischen Klassen in beliebigen Hierarchieebenen vorgeschlagen, wodurch eine Aggregation respektive Generalisierung von thematischen Klassen in größeren Hierarchieebenen erreicht werden kann. Dadurch wird sicher gestellt, dass die Segmente in größeren Skalenbereichen ebenfalls eine semantische Bedeutung haben. Für eine optimale Trennung der thematischen Klassen sowie im Zuge der Berücksichtigung der unterschiedlichen Segmentgrößen in den verschiedenen Skalen, sollen jeweils die optimalen Bildmerkmale verwendet werden. Ob hier prinzipiell ein generativer oder diskriminativer Ansatz verfolgt werden sollte, wird im folgenden Abschnitt diskutiert.

3 Wahrscheinlichkeitstheoretische graphenbasierte Bildmodellierung

Statistische graphenbasierte Modelle sind ein Werkzeug für das Lösen von komplexen und mit Unsicherheiten behafteten Problemstellungen. Ein komplexes System wird hier aus einfachen und elementaren Teilen zusammengesetzt, wobei Zusammenhänge und Interaktionen zwischen den definierten Variablen in Form von kontextuellen Bedingungen dargestellt werden. Die ursprüngliche Aufgabenstellung in dieser Arbeit ist die Klassifikation von Bildsegmenten und kann als inverses Problem aufgefasst werden, bei dem die „beste“ Realisierung der so genannten versteckten Variablen x (Klassenlabels) gegeben den beobachteten Variablen y (Bildaten) gesucht ist, wobei y in bestimmter Weise mit den versteckten Variablen interagiert. Da es sich um statistische Ansätze handelt, werden die Variablen x und y als Ereignisse der Zufallsvektoren X und Y aufgefasst. Es wird eine diskrete Menge von K Klassen angenommen, bei der die Wahrscheinlichkeit, dass eine Zufallsvariable X_i einen Wert x_i annimmt, mit $P(X_i = x_i)$ bezeichnet wird. Im Folgenden wird für diese und andere Wahrscheinlichkeiten eine Abkürzung der Form $P(X_i = x_i) = P(x_i)$ verwendet. Die gemeinsame Wahrscheinlichkeit einer Konfiguration von Zufallsvariablen x ist $P(X = x) = P(X_1 = x_1, \dots, X_n = x_n)$ und wird mit $P(x)$ abgekürzt. Ein einzelnes Element (Knoten) eines Graphen wird im Folgenden mit i oder auch x_i bezeichnet, wobei x_i gleichzeitig auch die Variable für die Klassenzugehörigkeit von Element i ist.

Generell können Verfahren der statistischen graphenbasierten Bildmodellierung in kausale und nicht kausale Modelle eingeteilt werden. In diesem Abschnitt erfolgt die detaillierte Darstellung von Methoden dieser beiden Gruppen, wobei jeweils sowohl planare als auch hierarchische Modelle behandelt werden.

3.1 Nichtkausale Modelle

Nichtkausale Modelle sind globale Wahrscheinlichkeitsverteilungen, welche bezüglich ungerichteter Graphen definiert sind. Für die Formulierung und Darstellung kontextueller Bedingungen werden Cliquespotentialiale $\Psi_c(x_c)$ verwendet:

$$P(x) \propto \prod_{c \in C} \Psi_c(x_c), \quad (3.1)$$

wobei C die Menge aller Cliques des Graphen ist. Eine Clique ist ein definierter Subgraph des ursprünglichen Graphen. Das Produkt (3.1) ist mit einer Normierungskonstante Z , welcher durch die Integration beziehungsweise Summation des Produktes über alle möglichen Konfigurationen x berechnet wird, proportional zu $P(x)$. Wie in Abschnitt 2 beschrieben wurde, sind planare MRFs üblicherweise generative ungerichtete Modelle. Dabei wird X als MRF bezüglich eines Nachbarschaftssystems \mathcal{N} bezeichnet, wenn die folgenden beiden Bedingungen

erfüllt sind

$$\begin{aligned} P(x) &> 0, \forall x \in X && \text{(Positivität)} \\ P(x_i | x_{x \setminus \{i\}}) &= P(x_i | x_{\mathcal{N}_i}) && \text{(Markov-Eigenschaft)} \end{aligned} \quad (3.2)$$

Der Ausdruck $x_{\mathcal{N}_i} = \{x_{i'} | i' \in \mathcal{N}_i\}$ beschreibt die Menge der Nachbarn von Bildelement i und $x \setminus \{i\}$ ist die Menge aller Knoten ohne Element i . Während die Positivität aus rechnerischen Gründen angenommen wird, beschreibt die Markov-Eigenschaft die lokalen Charakteristika des Zufallsfeldes. Eine mathematisch tragbare Darstellung von $P(x)$ wird durch das Hammersley-Clifford Theorem [Besag, 1974] ermöglicht. Diese so genannte Markov-Gibbs Äquivalenz beschreibt den Zusammenhang zwischen MRFs und Gibbs-Zufallsfeldern (GRFs) und besagt, dass diese beiden Felder äquivalent sind. Während ein MRF durch lokale Eigenschaften (Markov-Eigenschaften) definiert wird, wird ein GRF durch seine globalen Eigenschaften, der Gibbs-Verteilung

$$P(x) = Z^{-1} \exp(-U(x)) \quad (3.3)$$

charakterisiert, wobei $U(x) = \sum_{c \in C} V_c(x)$ die Gibbs-Energiefunktion, C die Menge aller Cliques, Z die Normalisierungskonstante und V_c das Cliquespotential für Clique c ist. Bei Klassifikationsproblemen wird für die Zufallsvariable X üblicherweise ein homogenes und isotropes MRF, wie das Ising-Modell (binäres Klassifikationsproblem) oder das Potts-Modell (Mehrklassenproblem) angenommen. Dies bedeutet, dass für X eine Gitterform angenommen wird und Interaktionen zwischen den beiden (eindimensionales Feld) oder den vier (zweidimensionales Feld) unmittelbaren Nachbarn erlaubt sind. Die Potentiale $V_c(x)$ sind dabei unabhängig von der relativen Position (Homogenität) sowie der Orientierung (Isotropie) der Cliques.

Auf der Grundlage des Hammersley-Clifford Theorems kann die Verteilung $P(x)$ nun durch die Definition zweckdienlicher Potentialfunktionen $V_c(x)$ spezifiziert werden. Mit Cliquespotentialen bezüglich maximal zwei Elementen (maximal paarweise Potentiale) hat die Energiefunktion die folgende Form

$$U(x|y) = \sum_{i \in x} V_1(y_i | x_i) + \sum_{i \in x} \sum_{i' \in \mathcal{N}_i} V_2(x_i, x_{i'}). \quad (3.4)$$

Bei einem Potts-Modell ist $V_2(x_i, x_{i'}) = -\beta \delta_k(x_i, x_{i'})$, mit der Kronecker-Delta Funktion

$$\delta_k(x_i, x_{i'}) \begin{cases} 1, & \text{if } x_i = x_{i'} \\ 0, & \text{if } x_i \neq x_{i'} \end{cases} \quad (3.5)$$

und V_1 sind die bedingten *likelihoods* $p(y_i | x_i)$ aus einer Bayes-Modellierung mit Gauß'schen Wahrscheinlichkeitsdichtefunktionen. Während V_1 demnach als Datenterm aufgefasst werden kann, bei dem lediglich das zentrale Element i respektive dessen Bilddaten y_i einfließen, ist V_2 der von den Bilddaten unabhängige Kontextterm.

3.1.1 Parameterschätzung

Die Schätzung eines Satzes von Modellparametern θ erfolgt durch die Optimierung eines Kriteriums, beispielsweise mittels ML, *expectation-maximization* (EM) oder Bayes'scher Schätzung. Sofern dazu bekannte Trainingsdaten für jede Klasse vorliegen, wird von einer überwachten

Schätzung gesprochen. Die ML-Schätzung erfolgt über die Maximierung der bedingten Wahrscheinlichkeit $P(x|\theta)$ (*likelihood* von θ) durch

$$\hat{\theta} = \arg \max_{\theta} P(x|\theta). \quad (3.6)$$

Bei dem oben behandelten Ising- beziehungsweise Potts-Modell kann die Schätzung der Parameter der angenommenen Gauß-Verteilungen mit den bekannten Gleichungen für Mittelwert μ und Kovarianzmatrix Σ , oder durch ein *clustering*, beispielsweise mittels EM [Dempster et al., 1977] erfolgen.

Der Modellparameter β (Gewicht des Kontextterms) wird in vielen Anwendungen manuell gewählt. Bei einer ML-Schätzung des Kontextparameters β muss prinzipiell die Normalisierungsfunktion Z der korrespondierenden Gibbs-Verteilung bestimmt werden. Die *likelihood*-Funktion in Gibbs-Form ist

$$P(x|\beta) = Z^{-1} \exp(-U(x|\beta)), \quad (3.7)$$

mit $U(x|\beta) = \sum_i \sum_{i' \in \mathcal{N}_i} V_2(x_i, x_{i'}, \beta)$ und

$$Z(\theta) = \sum_{x \in X} \exp(-U(x|\beta)), \quad (3.8)$$

wobei X die Menge aller möglichen Konfigurationen von Klassenzugehörigkeiten ist. Die Berechnung von (3.8) ist rechnerisch nicht tragbar, so dass Approximationen für die Lösung des Problems herangezogen werden müssen. Die Bestimmung der Normalisierungskonstante kann beispielsweise mit Hilfe der *pseudo-likelihood* (PL) [Besag, 1975], [Besag, 1977] umgangen werden. Die PL basiert auf den bedingten Wahrscheinlichkeiten $P(x_i|x_{\mathcal{N}_i}, \beta)$, wobei für jede lokale Energiefunktion U_i eine bekannte Nachbarschaft $x_{\mathcal{N}_i}$ angenommen wird. Die PL ergibt sich damit aus

$$PL(x|\beta) = \prod_{i \in x} P(x_i|x_{\mathcal{N}_i}, \beta), \quad (3.9)$$

wobei x alle Elemente des regulären Gitters sind. Daraus resultiert

$$PL(x|\beta) = \prod_i Z_i^{-1} \exp\left(-\sum_{i' \in \mathcal{N}_i} V_2(x_i, x_{i'}, \beta)\right), \quad (3.10)$$

wobei Z_i jeweils eine lokale Normalisierungskonstante ist. Auf diese Weise wird also die Bestimmung der globalen Normalisierungskonstante Z umgangen. Eine Maximierung der *log-pseudolikelihood* kann durch Lösen von

$$\frac{\partial \ln PL(x|\beta)}{\partial \beta} = 0 \quad (3.11)$$

beispielsweise mit dem Newton-Raphson Verfahren erfolgen [Cadez und Smyth, 1998].

3.1.2 Inferenz

Sofern die *a priori* Verteilung $P(x)$ sowie die Parameter der *likelihood* Funktionen $p(y|x)$ bekannt sind, kann das optimale Klassifikationsergebnis mit Hilfe der Bayes'schen Schätzung bestimmt werden. Die *maximum a posteriori* Schätzung (MAP) [Geman und Geman, 1984] ist ein sehr häufig angewendeter Spezialfall. Die MAP-Schätzung erfolgt mit Hilfe einer Maximierung der *a posteriori* Wahrscheinlichkeit

$$\hat{x} = \arg \max_x P(x|y). \quad (3.12)$$

Der einfachste Ansatz dafür ist mit dem *iterative conditional modes*-Algorithmus (ICM) [Besag, 1986] gegeben. Hier werden iterativ die lokalen bedingten Wahrscheinlichkeiten $P(x_i|y, x_{x \setminus \{i\}})$ maximiert. Es werden dafür zwei Annahmen getroffen. Zum einen werden die beobachteten Bilddaten y_1, \dots, y_n als bedingt unabhängig bei gegebenem x angenommen und für jedes x_i dieselben Dichtefunktionen $p(y_i|x_i)$ vorausgesetzt. Zum anderen wird wiederum die Markov-Eigenschaft, also die Konditionierung von x_i durch die lokalen Nachbarn \mathcal{N}_i , vorausgesetzt. Mit diesen beiden Annahmen und dem Bayes'schen Theorem folgt für die lokalen Wahrscheinlichkeiten

$$P(x_i|y, x_{x \setminus \{i\}}) \propto p(y_i|x_i) P(x_i|x_{\mathcal{N}_i}). \quad (3.13)$$

Die Maximierung von (3.13) ist äquivalent zu der iterativen Minimierung der Energiefunktion (3.4) durch

$$x_i^{k+1} \leftarrow \arg \min_{x_i} U(x_i|y_i, x_{\mathcal{N}_i}^k). \quad (3.14)$$

Eine häufig verwendete initiale Realisierung der Klassenlabels x^0 ist das Ergebnis einer ML-Schätzung $x^0 = \arg \max_x p(y|x)$. Die Iteration wird so lange fortgesetzt, bis ein Kriterium, beispielsweise die Unterschreitung einer minimalen Anzahl an Änderungen der Klassenzugehörigkeit pro Iteration oder der minimalen Änderung der Energiefunktion, erfüllt ist. Nach [Besag, 1986] ist mit diesem Verfahren immer Konvergenz garantiert.

Bei dem ICM-Algorithmus handelt es sich um ein lokales Verfahren, bei dem die Güte des Ergebnisses stark von der Initialisierung der Klassenlabels x^0 abhängt. Das *simulated annealing* (SA) [van Laarhoven und Aarts, 1987] ist hingegen ein globales statistisches Verfahren. Das Konzept von SA basiert auf der Simulation eines Prozesses der Abkühlung (Erhitzen und anschließendes Abkühlen einer physischen Substanz), mit dem Ziel eines niedrigen Energiegleichgewichts. Damit können komplexe kombinatorische Probleme gelöst werden. In [Geman und Geman, 1984] wird gezeigt, dass die MAP-Schätzung von MRFs durch dieses Verfahren erfolgen kann. Theoretisch wird mit SA das globale Minimum gefunden, jedoch wird für diesen Prozess unendlich viel Zeit benötigt. Für eine praktische Anwendung wird daher ein Kompromiss zwischen Laufzeit und Qualität der Lösung erforderlich [Hegarat-Masclé und Vidal-Madjar, 1996]. Ein praktischer Vergleich von SA mit dem ICM-Algorithmus in [Laferté et al., 2000] zeigt, dass für SA Laufzeiten mit einem Faktor von ca. 40 – 70 langsamer erreicht werden. Die Klassifikationsgüte ist jedoch mit SA bei einer Anwendung auf künstliche verrauschte Bilder signifikant besser.

3.2 Hierarchische nichtkausale Modelle

Nichtkausale hierarchische Modelle sind Wahrscheinlichkeitsverteilungen, die bezüglich ungeordneter hierarchischer Graphen definiert sind. Basierend auf den diskriminativen CRFs [Laf-

ferty et al., 2001] und DRFs [Kumar und Hebert, 2003a] wird in [Awasthi et al., 2007] ein hierarchischer Ansatz vorgestellt, bei dem für die Bildrepräsentation von Nahbereichsfotografien ein Quadtree verwendet wird (Abbildung 3.1). Paarweise Cliques werden durch die Knoten und Kanten des Baumes gebildet. Jeder innere Knoten wird mit einer versteckten Variable h assoziiert, wobei die Menge aller Knoten in einer Ebene H_n , mit den Levels $n = 1, \dots, N$, ist. Jeder Knoten der untersten versteckten Ebene (hier H_3) ist mit einem weiteren Knoten, der jeweils das gesuchte Klassenlabel x repräsentiert, verbunden. Die Bilddaten y werden in diesem Ansatz lediglich auf der feinsten Ebene des Quadrees betrachtet, wobei hier rechteckige Bildelemente der Größe 4×6 und 8×12 Pixel klassifiziert werden sollen.

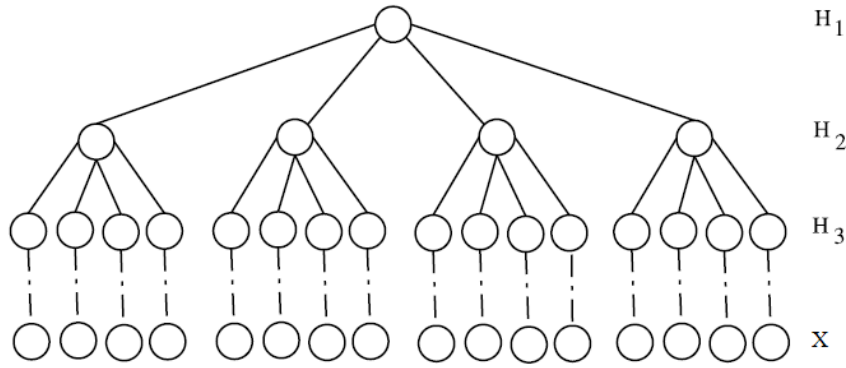


Abbildung 3.1: Quadtree-Struktur für ein TCRF. Quelle: [Awasthi et al., 2007].

Im Gegensatz zu einem generativen MRF handelt es sich hier um ein diskriminatives Verfahren, bei dem die *a posteriori* Wahrscheinlichkeit $P(x, h|y)$ direkt und ohne explizite Formulierung von Daten- und *a priori* Modell modelliert wird. Die gemeinsame bedingte Verteilung $P(x, h|y)$ lässt sich aufgrund der Baumstruktur in ein Produkt von Potentialfunktionen bezüglich der paarweisen Cliques zerlegen. Diese Potentialfunktionen sind wie folgt definiert.

Lokales Potential

Das lokale Potential drückt den Einfluss der beobachteten Bildinformation y auf die Klassenlabels x aus und beschreibt somit, wie wahrscheinlich ein Bildelement i ein Klassenlabel x_i bei gegebenen Beobachtungen y annimmt. Für jede Variable x_i hat das lokale Potential die Form $\exp(\Gamma_i(x_i, y))$ mit

$$\Gamma_i(x_i, y) = x_i^T W f_i(y). \quad (3.15)$$

Die Verwendung von beliebigen Bilddaten y verdeutlicht den Unterschied zu dem zuvor beschriebenen generativen Verfahren, bei dem an korrespondierender Stelle (Datenmodell) jeweils die entsprechende Bildinformation y_i des Bildelements einfließt. Es sei K die Anzahl der möglichen Klassen. Dann ist x_i ein binärer Vektor der Länge K , der an $K - 1$ Stellen eine 0 und an der Stelle der vorliegenden Klasse eine 1 hat. Des Weiteren ist $f_i(\cdot) = \{f_1, \dots, f_F\}$ eine für das i -te Element des Bildes berechnete Transformation der Bildinformation. Die Gewichtsmatrix W hat die Dimension $K \times F$ und repräsentiert den Zusammenhang zwischen $f_i(\cdot)$ und den Klassenlabels. Gleichung (3.15) entspricht einer linearen Regression. Prinzipiell

kann an dieser Stelle, äquivalent zu den generativen MRFs, ein beliebiger lokaler Klassifikator verwendet werden [He et al., 2006].

Kantenpotential

Die Kantenpotentiale repräsentieren die Interaktionen zwischen den versteckten Variablen in Ebene H_N (in Abbildung 3.1 Ebene H_3) und den Klassenlabels des Quadrees sowie zwischen den versteckten Variablen in den jeweils benachbarten Ebenen darüber. Dieses Potential drückt aus, wie wahrscheinlich ein versteckter Knoten i in Ebene n ein Klassenlabel x_i aufweist, wenn der Knoten j in Ebene $n + 1$ darunter ein bestimmtes Klassenlabel x_j hat. Die Potentiale für die inneren Knoten werden mit den Matrizen $\phi^{n,b}$ der Dimension $K \times K$ dargestellt. Da es sich um die Beziehung zwischen einem Knoten und seinen vier Kindknoten in der Baumstruktur handelt, gilt $b = 1, 2, 3, 4$. Damit ist das Potential zwischen einem inneren Knoten h_i und dessen Kindknoten h_j gegeben durch

$$\Psi_{i,j}(h_i, h_j) = \exp\left(h_i^T \phi^{n,b} h_j\right). \quad (3.16)$$

Das Kantenpotential zwischen den Knoten der untersten versteckten Ebene $h_i \in H_N$ und den gesuchten Klassenlabels x ist definiert durch

$$\Psi_{i,j}(h_i, x_j) = \exp\left(h_i^T \phi^N x_j\right). \quad (3.17)$$

Wahrscheinlichkeitsverteilung

Die gemeinsame Wahrscheinlichkeitsverteilung für dieses Modell ist ein Produkt der obigen Potentiale

$$\begin{aligned} P(x, h|y) \propto \exp & \left(\sum_{x_i \in x} x_i^T W f_i(y) \right. \\ & + \sum_{x_j \in x} \sum_{h_i \in H_N} h_i^T \phi^N x_j \\ & \left. + \sum_{n=1}^{N-1} \sum_{h_i \in H_n} \sum_{h_j \in H_{n+1}} h_i^T \phi^{n,b} h_j \right). \end{aligned} \quad (3.18)$$

3.2.1 Parameterschätzung

Aufgrund des schleifenlosen Graphen bei eindimensionalen CRFs können die Parameter $\Theta = \{W, \phi\}$ durch eine exakte ML-Schätzung erfolgen. Die Berechnung der Normalisierungskonstante Z kann bei schleifenfreien Graphen durch verschiedene effiziente Verfahren, wie beispielsweise *conjugate gradient* [Wallach, 2002] und *gradient boosting* [Dietterich et al., 2004] erfolgen. Die Quadtree-Struktur ist ebenfalls frei von Schleifen, jedoch ist aufgrund der versteckten Variablen h keine exakte Parameterschätzung möglich. Die Maximierung der bedingten *log-likelihood* der Trainingsdaten ist äquivalent zu einer Minimierung der Kullback-Leibler-Divergenz (KL) zwischen der Modellverteilung $P(x|y, \Theta)$ und der empirischen Verteilung $Q(x|y)$, welche durch die Trainingsdaten definiert ist [Awasthi et al., 2007]. Die Berechnung der dafür erforderlichen Erwartungswerte bezüglich der Modellverteilung unter Verwendung der üblichen *Makov chain Monte Carlo* (MCMC) Methoden dauert sehr lange wobei die

geschätzten Gradienten zu einer hohen Varianz tendieren [Awasthi et al., 2007]. Das *contrastive divergence* (CD) Verfahren [Hinton, 2002] ist ein approximatives Verfahren, bei dem diesen Nachteilen durch die Minimierung einer alternativen Zielfunktion entgegen gewirkt wird. Diese Zielfunktion wird *contrastive divergence* genannt und ist gegeben mit

$$D = Q^0 \|P - Q^1\|P \quad (3.19)$$

wobei $Q \|P = \sum_{x \in X} Q(x) \log \frac{Q(x)}{P(x)}$ die KL-Divergenz zwischen Q und P ist. Q^1 ist eine Verteilung, die durch eine so genannte Ein-Schritt-Rekonstruktion der Trainingsdaten definiert ist [Hinton, 2002]. Die Verbesserung der Gewichte im Training wird bestimmt aus

$$\Delta w_{ij} = -\eta \frac{\partial D}{\partial w_{ij}}, \quad (3.20)$$

wobei η die Lernrate ist. Die Ein-Schritt-Rekonstruktion stellt eine einzelne Iteration der MCMC-Abtastmethode dar. Laut [Hinton, 2002] ist dieser Schritt für eine Konvergenz gegen die ML-Schätzung ausreichend und wird wie folgt bestimmt:

1. Wähle eine Konfiguration x^0 entsprechend der Verteilung Q^0 .
2. Berechne für jede Potentialfunktion die *a posteriori* Wahrscheinlichkeitsverteilung über die versteckten Variablen gegeben x^0 .
3. Wähle auf der Grundlage dieser Verteilungen die versteckten Variablen (*sampling*).
4. Berechne die bedingte Verteilung über alle sichtbaren Variablen auf der Grundlage der gegebenen versteckten Variablen.
5. Wähle für jede sichtbare Variable auf der Grundlage der bedingten Verteilung einen Wert. Diese Werte bilden den rekonstruierten Vektor x^1 .

Sei x^1 die Ein-Schritt-Rekonstruktion der Trainingsdaten x^0 , so ergeben sich die folgenden Verbesserungsgleichungen der Gewichte in einer Iteration

$$\Delta w_{uv} = \eta \left[\sum_{x_i \in x^0} x_i^u f_i^v(y) - \sum_{x_i \in x^1} x_i^u f_i^v(y) \right] \quad (3.21)$$

$$\begin{aligned} \Delta \phi_{uv}^N = \eta & \left[\sum_{x_i \in x^0} \sum_{h_i \in H_N} E_{P(h_i|x^0)} [h_i^u x_j^v] \right. \\ & \left. - \sum_{x_i \in x^1} \sum_{h_i \in H_N} E_{P(h_i|x^1)} [h_i^u x_j^v] \right] \end{aligned} \quad (3.22)$$

$$\begin{aligned} \Delta \phi_{uv}^{n,b} = \eta & \left[\sum_{h_i \in H_n} \sum_{h_j \in H_{n+1}} E_{P(h_i, h_j|x^0)} [h_i^u h_j^v] \right. \\ & \left. - \sum_{h_i \in H_n} \sum_{h_j \in H_{n+1}} E_{P(h_i, h_j|x^1)} [h_i^u h_j^v] \right] \end{aligned} \quad (3.23)$$

Durch die Bildung des Quotienten bei der KL-Divergenz kürzt sich die Normalisierungskonstante komplett heraus. Das Verfahren wurde auch in [Gagrani, 2002] erfolgreich für die Parameterschätzung verwendet.

3.2.2 Inferenz

Die gesuchten Variablen x werden in [Awasthi et al., 2007] durch die *marginal posterior mode* (MPM) Inferenz [Peleg et al., 1984] geschätzt. Das Kriterium bei der MPM-Inferenz führt zu einer Minimierung der globalen Anzahl an falsch klassifizierten Bildelementen und lautet [Marroquin et al., 1987]

$$\hat{x}_i = \arg \max_{x_i} P(x_i|y). \quad (3.24)$$

Dabei bedeutet die Verwendung von y , dass nicht nur die Bildinformation eines Bildelements, sondern alle Informationen einfließen. Aufgrund der Baumstruktur des Graphen können die bedingten Wahrscheinlichkeiten $P(x_i|y)$ mittels *belief propagation* (PB) [Pearl, 1988] exakt berechnet werden. Hier werden Nachrichten in Form von realen Zahlen über die Kanten des Graphen zwischen den Knoten ausgetauscht. Eine *a posteriori* Randdichte $P(x_i|y)$ ergibt sich aus dem Produkt aller eingehenden Nachrichten der benachbarten Knoten. In Abschnitt 4.5 wird das Verfahren an einem Beispiel näher beschrieben.

3.3 Kausale Modelle

Kausale Modelle sind globale und auf gerichtete Graphen definierte Wahrscheinlichkeitsverteilungen. Die kontextuellen Beziehungen zwischen den Knoten des Graphen werden mit Hilfe von Übergangswahrscheinlichkeiten formuliert, wobei angenommen wird, dass die beobachteten Zufallsvariablen durch einen kausalen Prozess generiert wurden. Wird ein schleifenfreier (azyklischer) Unabhängigkeitsgraph angenommen, ergibt sich die gemeinsame Verteilung der Variablen x zu

$$P(x) = \prod_i P(x_i|pa_i), \quad (3.25)$$

wobei pa_i die Elternknoten von Knoten i sind. Somit sind alle Knoten, die sich nicht in einer Eltern-Kind Beziehung befinden, bedingt unabhängig gegeben pa_i . Gerichtete azyklische Graphen sind unter dem Namen Bayes'sche Netze oder auch *belief net* (BN) bekannt. Ein

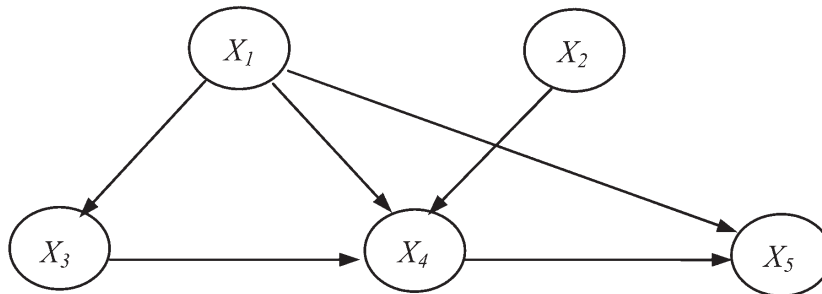


Abbildung 3.2: Einfache Struktur für ein Bayes'sches Netz. Quelle: [Ouyang et al., 2006].

einfaches Beispiel für ein Bayes'sches Netz ist in Abbildung 3.2 dargestellt. Die entsprechende

gemeinsame Wahrscheinlichkeit nach (3.25) ist

$$P(X_1, X_2, X_3, X_4, X_5) = P(X_1)P(X_2)P(X_{X_3}|X_1)P(X_4|X_1, X_2, X_3)P(X_5|X_1, X_4). \quad (3.26)$$

Die bedingten Wahrscheinlichkeiten werden für jeden Knoten in Form einer Matrix mittels bedingter Wahrscheinlichkeitstabellen (*conditional probability tables*, CPT) bereitgestellt. Bayes'sche Netze werden unter anderem für die Klassifikation von Fernerkundungsdaten verwendet. Es ist dabei zu beachten, dass die Graphenstruktur hier nicht das Bildmodell repräsentiert, sondern vielmehr den kausalen Prozess, der die Generierung der Bilddaten modelliert.

3.3.1 Parameterschätzung

In [Ouyang et al., 2006] erfolgt die Klassifikation der Landbedeckung von Landsat ETM+ Satellitenbildern mittels eines Bayes'schen Multinetz-Klassifikators. Dabei wird für jede definierte Landbedeckungsklasse ein spezifisches Bayes'sches Netz angenommen, so dass die statistischen Relationen zwischen den Bildmerkmalen für jede Klasse unterschiedlich sein können. Für das Lernen der optimalen Graphenstrukturen wird ein auf der Messung von Abhängigkeitsrelationen basierender Algorithmus [Cheng et al., 1997] angewendet, der an dieser Stelle nicht behandelt werden soll.

Ziel der Parameterschätzung ist die Besetzung der CPTs aller Knoten eines Graphen G gegeben den Trainingsdaten. Dies kann mit einer ML-Schätzung [Murphy, 2002] erfolgen. Gegeben ist ein Trainingsdatensatz x^T , welcher $i = 1, \dots, M$ Ereignisse (Bildelemente) beinhaltet. Die Ereignisse $x_1^T, x_2^T, \dots, x_M^T$ werden als unabhängig angenommen, so dass die *log-likelihood* des Trainingsdatensatzes mit dem folgenden Ausdruck gegeben ist

$$L = \log \prod_{m=1}^M P(x_m^T|G) = \sum_{i=1}^n \sum_{m=1}^M \log P(x_i|pa_i, x_m^T), \quad (3.27)$$

wobei $i = 1, \dots, n$ die Knoten des Graphen und $m = 1, \dots, M$ die Trainingsdaten indiziert und pa_i die Menge der Elternknoten von x_i ist. Jede Variable x_i kann einen Zustand aus $k = 1, \dots, r_i$ verschiedenen Werten annehmen und pa_i hat q_i verschiedene Konfigurationen. Die Matrix mit den gesuchten Übergangswahrscheinlichkeiten ist durch

$$\theta_{ijk} = p(x_i^k|pa_i^j) \quad (3.28)$$

definiert, wobei $1 \leq i \leq n$ (Knoten), $1 \leq j \leq q_i$ (Konfiguration der Elternknoten) und $1 \leq k \leq r_i$ (mögliche Zustände in einem Knoten) gilt. Die *log-likelihood* L ergibt sich damit zu

$$\begin{aligned} L &= \sum_{i=1}^n \sum_{m=1}^M \log \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} \theta_{ijk}^{I_{ijkm}} \\ &= \sum_{i=1}^n \sum_{m=1}^M \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} I_{ijkm} \log \theta_{ijk} \\ &= \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} N_{ijk} \log \theta_{ijk}, \end{aligned} \quad (3.29)$$

mit

$$I_{ijkm} = \begin{cases} 1 & \text{wenn } x_i = x_i^k, pa_i = pa_i^j \text{ in Ereignis } D_m \text{ auftritt} \\ 0 & \text{sonst} \end{cases} \quad (3.30)$$

$N_{ijk} = \sum_{m=1}^M I_{ijkm}$ ist die Anzahl des Auftretens von Ereignis $(x_i = x_i^k, pa_i = pa_i^j)$ im Trainingsdatensatz x^T . Der ML-Schätzer ist somit gegeben durch

$$\hat{\theta}_{ijk} = \frac{N_{ijk}}{\sum_{k'=1}^{r_i} N_{ijk'}}. \quad (3.31)$$

3.3.2 Inferenz

Bei gerichteten azyklischen Graphen liegt generell eine Struktur vor, die eine effiziente Marginalisierung ermöglicht. Dafür kann, wie auch in den bisher vorgestellten Modellen, das Nachrichten-Schema nach [Pearl, 1988] (*message-passing*) angewendet werden. Für die Aktualisierung der Wahrscheinlichkeiten in einem Knoten x_i , mit einer Menge von Elternknoten pa_i sowie Kindknoten ch_i , werden Botschaften sowohl in Pfeilrichtung (*top-down*) als auch in entgegengesetzter Pfeilrichtung (*bottom-up*) propagiert. Das Verfahren wird detailliert in Abschnitt 4.5 beschrieben.

3.4 Hierarchische kausale Modelle

In hierarchischen kausalen Bildmodellen werden die miteinander interagierenden Knoten des Graphen in verschiedene Hierarchieebenen gruppiert. Die unterste dieser Ebenen repräsentiert die Menge der kleinsten Bildelemente und damit die gesuchten Klassenlabels x . Eine häufig verwendete Graphenstruktur ist ein kausaler Quadtree. Äquivalent zu einem klassischen Bayes'schen Netz wird hier die Markov-Eigenschaft angenommen. Um Unabhängigkeitsannahmen bezüglich des Datenmodells zu lockern, wird bei einem Multiskalen-Zufallsfeld (*multiscale random field*, MSRF) [Kumar und Hebert, 2003b] an Stelle eines herkömmlichen Quadtrees eine etwas komplexere Graphenstruktur verwendet (Abbildung 3.3). Die gesuchten Klassen-

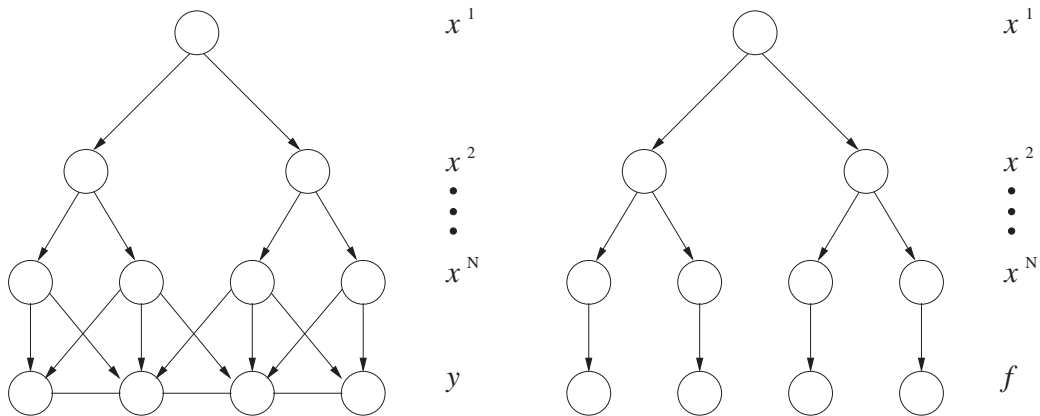


Abbildung 3.3: Dyadische Darstellung der Graphenstruktur eines MSRF. Links: originale Struktur, Rechts: approximierte Struktur. Quelle: [Kumar, 2005].

labels in den N Levels des Baumes werden hier mit $x = \{x^1, x^2, \dots, x^N\}$ bezeichnet, wobei

x^n die Menge aller Knoten in Level n ist. In diesem mehrskaligen Zufallsfeld impliziert die Markov-Eigenschaft bezüglich der Skalen

$$P(x^n|x^1, \dots, x^{n-1}) = P(x^n|x^{n-1}), \text{ für } n = 2, \dots, N. \quad (3.32)$$

Des Weiteren wird für gerichtete Graphen eine bedingte Unabhängigkeit der Klassenlabels in einem Skalenbereich angenommen, so dass

$$P(x^n|x^{n-1}) = \prod_{i \in x^n} P(x_i^n|x_i^{n-1}) \quad (3.33)$$

gilt. Wie in Abbildung 3.3 links angedeutet, werden die beobachteten Daten y nicht als bedingt unabhängig gegeben den Klassenlabels angesehen. Es wird vielmehr eine Abhängigkeit von den Klassenlabels und den Bildinformationen der Elemente in der Nachbarschaft angenommen. Aus Gründen der rechnerischen Tragbarkeit wird für das Datenmodell wiederum eine faktorisierte Form

$$P(y|x^N) \approx \prod_{i \in x^N} P(y_i|y_{\mathcal{N}_i}, x_i^N) \quad (3.34)$$

angenommen, wobei \mathcal{N}_i die Nachbarschaft des Elements i ist, und $y_{\mathcal{N}_i} = \{y_{i'}|i' \in \mathcal{N}_i\}$. Diese Approximation ist äquivalent zu der *pseudo-likelihood* Faktorisierung in Abschnitt 3.1.1. Es wird dabei vernachlässigt, dass die beobachteten Daten y_i eines jeden Elements auch von den Klassenlabels der Elemente in dessen Nachbarschaft abhängen. Die gemeinsame Verteilung in diesem Modell lautet mit den oben getroffenen Annahmen

$$P(x, y) = P(x^r) \prod_{i \in x} P(x_i|pa_i) \prod_{i \in x^N} P(y_i|y_{\mathcal{N}_i}, x_i^N), \quad (3.35)$$

wobei x die Menge aller Knoten des Baumes ausgenommen der Wurzel x^r ist. Entsprechend der rechten Seite in Abbildung 3.3 werden die bedingten Wahrscheinlichkeiten $P(y_i|y_{\mathcal{N}_i}, x_i^N)$ durch $P(f_i|x_i^N)$ approximiert, wobei f_i ein Merkmalsvektor ist, der die Abhängigkeiten der Bilddaten in Element i zu seinen Nachbarelementen ausdrückt. In [Kumar und Hebert, 2003b] wird für die Erkennung von künstlichen Strukturen in Bildern ein mehrskaliger Merkmalsvektor für f_i definiert, der Bildmerkmale eines Bildelements bezüglich einer variierenden Fenstergröße um dieses Bildelement repräsentiert. Für die Parametrisierung des Datenmodells werden *Gaussian mixture* Modelle (GMM) verwendet.

3.4.1 Parameterschätzung

Die Parameterschätzung für das generative MSRF erfolgt sequentiell. Für die Schätzung der Parameter θ_d des GMMs (Gewichte, Mittelwerte und Kovarianzmatrizen) wird ein EM-Algorithmus verwendet. Die Parameter des *a priori* Modells θ_p sind Klassenübergangswahrscheinlichkeiten zwischen Knoten benachbarter Skalen. Eine Übergangswahrscheinlichkeit θ_{ikl} eines Knotens i ist definiert als $\theta_{ikl} = P(x_i = k|pa_i = l)$ mit der Bedingung $\sum_l \theta_{ikl} = 1$ und $k, l = 0, \dots, K$. Sie drückt die lokale bedingte Verteilung in Knoten i gegeben der Klasse l des Elternknotens in der darüber liegenden Ebene aus. Die Schätzung der Übergangswahrscheinlichkeiten erfolgt durch die Anwendung eines ML-Ansatzes [Feng et al., 2002], wonach die Wahrscheinlichkeit der Trainingsdaten durch

$$\hat{\theta}_p = \arg \max_{\theta_p} \prod_{m=1}^M P(x^{N^m}, y^m|\theta_p, \theta_d) \quad (3.36)$$

maximiert wird. Mit $m = 1, \dots, M$ werden die Trainingsbilder indiziert. Die Annahme fester Parameter θ_d für das Datenmodell führt zu dem Ausdruck

$$\hat{\theta}_p = \arg \max_{\theta_p} \prod_{m=1}^M P(x^{Nm} | \theta_p). \quad (3.37)$$

Im Zuge des EM-Algorithmus werden alle Klassenlabels der Knoten des MSRFs von der Wurzel bis zum Level $N - 1$ als versteckte Variablen $x_h = \{x \setminus x^N\}$ angesehen. Der E-Schritt kann als Konstruktion einer lokalen unteren Schranke der *a posteriori* Verteilung gesehen werden [Neal und Hinton, 1998]. Diese wird bei der aktuellen Schätzung der Modellparameter θ'_p durch den folgenden Erwartungswert berechnet:

$$Q(\theta_p, \theta'_p) = \sum_{m=1}^M E_{(x_h^m | \theta'_p)} [\log P(x_h^m, x^{Nm} | \theta_p)]. \quad (3.38)$$

Im M-Schritt werden die neuen Werte der Parameter durch die Maximierung der Schranke berechnet

$$\theta_{ikl} = \frac{\sum_{m=1}^M P(x_i^m = k, pa_i^m = l | x^{Nm}, \theta'_p)}{\sum_{m=1}^M \sum_{k'} P(x_i^m = k', pa_i^m = l | x^{Nm}, \theta'_p)}. \quad (3.39)$$

Dabei ist die Schätzung der *a posteriori* Wahrscheinlichkeiten der Eltern-Kind Paare durch

$$P(x_i^m = k, pa_i^m = l | x^{Nm}, \theta'_p) = \frac{\lambda(x_i^m = k) \theta'_{ikl} \pi(pa_i^m = l)}{\sum_{l'} \lambda(pa_i^m = l') \pi(pa_i^m = l')} \prod_{u \in U(x_i)} \lambda_u(pa_i^m = l) \quad (3.40)$$

erforderlich, wobei $\lambda(\cdot)$ und $\pi(\cdot)$ nach [Pearl, 1988] definiert sind (siehe Abschnitt 4.5) und $U(x_i)$ die Menge aller Geschwisterknoten von x_i ist. Aufgrund begrenzt vorhandener Trainingsdaten wird für jedes Level n üblicherweise die gleiche Matrix der Übergangswahrscheinlichkeiten θ_{nkl} angenommen.

3.4.2 Inferenz

In [Kumar und Hebert, 2003b] werden die bekannten Schätzer MAP und MPM angewendet, wobei wiederum das *message passing* Schema von Pearl angewendet wird. Eine detaillierte Darstellung der Inferenzmethoden erfolgt in Abschnitt 4.5.

3.5 Generativ versus diskriminativ

In den letzten Jahren wurden verstärkt diskriminative Verfahren vorgestellt. Diese stellen eine Alternative zu den klassischen generativen Verfahren dar. Bei generativen Verfahren wird die gemeinsame Verteilung $P(x, y) = p(y|x)P(x)$ der gesuchten Klassenlabels x und der Bildmerkmale y modelliert. Dabei werden die gesuchten Zufallsvariablen X als MRF angesehen und $p(y|x)$ explizit durch ein Datenmodell beschrieben. In diskriminativen Ansätzen wird durch die direkte Modellierung der *a posteriori* Wahrscheinlichkeit $P(x|y)$ die explizite Formulierung eines Datenmodells vermieden. Dies kann in den Fällen vorteilhaft sein, wo die zu Grunde liegende generative Struktur sehr kompliziert, die gesuchte *a posteriori* Verteilung hingegen einfach ist [Feng et al., 2002]. Im diskriminativen Ansatz von [Awasthi et al., 2007]

entspricht die Formulierung des Zusammenhangs von Bilddaten und der gesuchten Klassenzugehörigkeiten der Bildelemente einer linearen Regression. Auch wenn hier kein zusätzlicher Aufwand bezüglich der Modellierung der Bilddaten betrieben wird, muss bedacht werden, dass semantische Klassen in einem Merkmalsraum nicht immer linear trennbar sind und in diesen Fällen keine zufriedenstellenden Ergebnisse zu erwarten sind. Äquivalent zu den generativen Verfahren können jedoch auch bei diskriminativen Verfahren beliebige und separat trainierte Klassifikatoren verwendet werden.

Um eine rechnerische Tragbarkeit bei Parameterschätzung und Inferenz zu gewährleisten, werden in generativen Verfahren vereinfachende Annahmen bezüglich des Datenmodells (bedingte Unabhängigkeit der Beobachtungen gegeben den Klassenlabels) getroffen. Des Weiteren sind die paarweisen Potentiale gänzlich unabhängig von den Bilddaten. Bei diskriminativen Verfahren werden solche Unabhängigkeitsannahmen nicht getroffen, wodurch die Verwendung von überlappenden Bildmerkmalen (beispielsweise in unterschiedlichen Skalen) ermöglicht wird [Hoberg und Rottensteiner, 2010]. Des Weiteren ist die Formulierung der paarweisen Potentiale im Gegensatz zu generativen Ansätzen als eine Funktion sowohl von den Klassenlabels als auch von den Bilddaten möglich. Die bedingte Verteilung der Klassenlabels x gegeben den Beobachtungen y ist für den DRF-Ansatz nach [Kumar und Hebert, 2003a] demnach gegeben mit

$$P(x|y) = \frac{1}{Z} \exp \left(\sum_{i \in x} A_i(x_i, y) + \sum_{i \in x} \sum_{j \in \mathcal{N}_i} I_{ij}(x_i, x_j, y) \right), \quad (3.41)$$

wobei $A(\cdot)$ als Assoziationspotential und $I(\cdot)$ als Interaktionspotential bezeichnet wird. Diese beiden Potentiale sind jeweils abhängig von allen Bilddaten y , wohingegen in klassischen MRFs das Assoziationspotential lediglich von den Bilddaten y_i des jeweils betrachteten Bildelements i und das paarweise Interaktionspotential gänzlich unabhängig von den Bilddaten ist.

Ein genereller Vorteil bei generativen Verfahren ist die Möglichkeit des Einbezugs unbekannter Daten bei der Parameterschätzung für das Datenmodell durch so genannte *semi supervised* Verfahren. Im Bereich der Objektdetektion, wo mitunter hohe Variabilitäten bei den Eingangswerten auftreten, können durch eine Überlagerung von verfügbaren Trainingsdaten und einer größeren Menge unbekannter Daten ausreichend viele Trainingsdaten bereitgestellt werden [Lasserre und Bishop, 2007].

3.6 Diskussion

Sowohl für gerichtete als auch für ungerichtete hierarchische Modelle existieren generative und diskriminative Verfahren. Der Hauptvorteil diskriminativer Verfahren ist mit der Vermeidung von Unabhängigkeitsannahmen gegeben. So können beispielsweise für die Erkennung künstlicher Strukturen in Fotografien [Awasthi et al., 2007] sowie in hochauflösenden optischen Satellitenbildern [Hoberg und Rottensteiner, 2010] überlappende Bildmerkmale aus mehreren Skalen verwendet werden. Dieser Vorteil wird vereinzelt hervorgehoben, ohne tatsächlich davon Gebrauch zu machen (beispielsweise in [Richard et al., 2004]). In [Kumar und Hebert, 2003a], [Awasthi et al., 2007] und [Hoberg und Rottensteiner, 2010] wird für die Berechnung von Multiskalen-Merkmalvektoren das jeweilige Bildelement selbst, sowie zwei größere Ausschnitte mit diesem im Zentrum einbezogen. In [Kumar und Hebert, 2003a] und [Hoberg und Rottensteiner, 2010] handelt es sich dabei um ein planares Modell und in [Awasthi et al., 2007] um ein hierarchisches Modell, bei dem diese Merkmalsvektoren bezüglich der feinsten Ebene der jeweiligen Graphenstruktur berechnet werden. Bezüglich der Aufgabenstellung in

dieser Arbeit ist jedoch nicht *a priori* garantiert, dass die Verwendung solcher Merkmale einen Mehrwert bringt, da die thematische Aufgabenstellung, die gegebenen Bildcharakteristika sowie die Größe der Segmente vorab nicht bekannt sind. Des Weiteren steht der Verwendung von Multiskalen-Merkmalvektoren bezüglich der feinsten Skala dieser Modelle die freie Verwendung von Bildinformation auf mehreren Skalen des Quadtree-Modells gegenüber.

Ein weiterer Aspekt bei den hier angeführten Publikationen ist die meist sehr klar definierte Aufgabenstellung. Ein festes Anwendungsgebiet ermöglicht *a priori* die Definition einer Menge von problemrelevanten (Multiskalen-) Merkmalen, die aus der Bildinformation der Bildelemente berechnet werden können. Dies ermöglicht eine simultane Schätzung aller Modellparameter in diskriminativen Ansätzen, ohne die tatsächliche Relevanz einzelner Merkmale berücksichtigen zu müssen. Um in dieser Arbeit die gewünschte übertragbare Anwendbarkeit der Methode zu gewährleisten, soll eine große Menge initialer Merkmale bereitgestellt werden, aus der die jeweils relevanten Merkmale selektiert werden müssen, um redundante und irrelevante Merkmale auszuschließen, die Dimensionalität des Klassifikationsproblems zu reduzieren und die Berechnung aller Merkmale für alle Bildelemente zu vermeiden. Eine simultane Schätzung aller Modellparameter ist hier demnach nicht ohne großen Aufwand möglich. Die zudem sehr begrenzte Menge an Trainingsdaten ist ebenfalls ein Grund für eine sequentielle Vorgehensweise in dieser Arbeit.

Als weiterer Vorteil diskriminativer Verfahren kann die Möglichkeit der Formulierung der kontextuellen Beziehungen in Abhängigkeit der Bilddaten angeführt werden. In [Kumar und Hebert, 2003a] und [Kumar, 2005] wird beispielsweise ein datenabhängiges Interaktionspotential $I(\cdot)$ (Gleichung 3.41) eingeführt. Im Vergleich zu dem Kontextterm in einem generativen MRF kann dadurch beispielsweise eine Glättung abgeschwächt werden, wenn die Daten der jeweils betrachteten Bildelemente ausreichend unterschiedlich sind. Vor dem Hintergrund eines hierarchischen Modells, bei dem die benachbarten Bildelemente in unterschiedlichen Hierarchieebenen liegen und in diesen auch unterschiedliche Bildmerkmale und Klassen zugelassen sind (vergleiche Abschnitt 2.3), erscheint diese Vorgehensweise nicht sinnvoll. Generell wird die Aufgabe der Definition von zweckdienlichen Interaktions- und Assoziationspotentialen in diskriminativen Ansätzen und vor allem im Rahmen eines allgemein formulierten Ansatzes vom Autor als sehr schwierig eingeschätzt.

Die obigen Überlegungen und Ausführungen zeigen, dass sich die Anwendung eines diskriminativen Verfahrens in dieser Arbeit nicht eignet. Die Begründung dafür liegt in den Rahmenbedingungen dieser Arbeit. Zum einen verhindert das initiale Fehlen von Trainingsdaten, verbunden mit einer Minimierung der manuellen Schritte, wie beispielsweise der Angabe von Trainingsdaten, eine simultane Schätzung aller Modellparameter, wie sie in den meisten diskriminativen Ansätzen erfolgt. Es ist vielmehr zu erwarten, dass für die Schätzung der Gesamtmenge an Modellparametern nicht ausreichend repräsentative Trainingsdaten vorhanden sind. Daraus kann abgeleitet werden, dass das Rahmenwerk auf einer sequentiellen Prozessierung beruhen sollte, bei der die Parameterschätzung für Daten- und *a priori* Modell getrennt voneinander erfolgen kann und Zwischenergebnisse als Grundlage für weitere Prozessierungsschritte herangezogen werden.

Zum anderen ist aus Gründen der Flexibilität und Schnelligkeit der Methode eine Selektion von problemrelevanten Merkmalen erforderlich, was wiederum für eine sequentielle Vorgehensweise spricht. Für diskriminative Ansätze erfolgt dies aufgrund eines klar definierten Anwendungsgebiets oft nicht. Die Anwendung einer Merkmalsextraktion, beispielsweise der Hauptkomponentenanalyse, würde wiederum eine aufwändige Berechnung aller Bildmerkmale für alle Bildelemente bedeuten. Ein generativer Ansatz, bei dem die explizite Modellierung

der beobachteten Bilddaten durch ein Datenmodell (beispielsweise durch ein GMM) erfolgt, ermöglicht hingegen eine vorgeschaltete Merkmalsselektion. Der diskriminative Ansatz in [Awasthi et al., 2007] verdeutlicht zudem, dass auch der jeweilige lokale Klassifikator, wie beispielsweise eine lineare Regression (LR) oder eine *support vector machine* (SVM) keine Garantie für gute oder bessere Ergebnisse sein muss, denn eine lineare Trennbarkeit der Klassen im Merkmalsraum ist nicht zwingend gegeben (LR) und die Wahl einer adäquaten Kernelfunktion ist sehr schwierig (SVM).

4 Hybrides Markov-Modell

Das hybride¹ Markov-Modell setzt sich aus einem gerichteten hierarchischen und einem ungerichteten planaren Modell zusammen und basiert auf den Arbeiten von [Pérez et al., 2000] und [Laferté et al., 2000]. Das hierarchische Modell basiert auf einer Quadtree-Baumstruktur, wodurch hierarchische Kontextinformation sowie Bilddaten aus unterschiedlichen Abstraktionsbeziehungsweise Quadtree-Ebenen in die Klassifikation einfließen. Die starre Struktur kann allerdings zu Blockstrukturen im Inferenzergebnis führen (vergleiche Abschnitt 2.2.2). Um diesen Effekt zu verringern sowie das Ergebnis der hierarchischen Inferenz zu verbessern, wird das Quadtree-Modell in [Wilson und Li, 2002] um den Einbezug von räumlichem Kontext in jeder Hierarchieebene erweitert, so dass jeder innere Knoten nicht nur mit seinem Eltern- sowie seinen vier Kindknoten, sondern auch mit seinen vier unmittelbaren Nachbarn innerhalb seiner Hierarchieebene verbunden ist. Daraus resultiert wiederum ein zyklischer Graph und somit eine iterative und approximative Inferenz. Um die gewinnbringende Eigenschaft der nicht iterativen und exakten Inferenz durch den Quadtree zu erhalten, wird in [Feng et al., 2002] und [Martinis et al., 2011] der zusätzliche Einbezug von räumlichem Kontext nach der hierarchischen Inferenz vorgeschlagen. Dieser Ansatz entspricht der zuvor diskutierten sequentiellen Prozessierung und soll daher in dieser Arbeit angewendet werden, wobei neben einer starren Anwendung beider Modelle nacheinander diskutiert werden soll, wie die Informationen aus den Modellen miteinander kombiniert werden könnten. Im Folgenden werden die statistischen Eigenschaften sowie die getroffenen Annahmen im hybriden Modell beschrieben und auf dieser Grundlage die Methoden für die Parameterschätzung und die Inferenz dargestellt.

4.1 Statistische Eigenschaften des hybriden Modells

Im hierarchischen Modell wird jeder Knoten i des Quadrees mit einem Klassenlabel x_i sowie mit einem Vektor y_i , welcher die Bildmerkmale (Beobachtungen) repräsentiert, assoziiert (vergleiche Abbildung 4.1). Ein Quadtree hat $n = 1, \dots, N$ Level, die jeweils eine Menge Knoten x^n repräsentiert. Die Wurzel (*root*) des Baumes ist $x^0 = x_r$. Ein innerer Knoten x_i in Level n hat einen Elternknoten pa_i im Level $n - 1$ und eine Menge ch_i von Kindknoten in Level $n + 1$. Die folgenden statistischen Annahmen werden in diesem Modell getroffen.

1. Markov-Eigenschaft bezüglich der Skalen: Die Menge der Knoten $(x^n)_{n=0}^N$ des Quadrees ist eine *top-down* Markov-Kette erster Ordnung, so dass

$$P(x^n | x^{n-1}, \dots, x^0) = P(x^n | x^{n-1}), \quad \forall n \in \{1, \dots, N\}. \quad (4.1)$$

2. Die Komponenten x^n einer Ebene des Quadrees werden als räumlich unabhängig bei gegebenen Knoten x^{n-1} angenommen. Mit der obigen Annahme einer Markov-Kette

¹Die Bezeichnung *hybrid* für diese Modellform wird erstmals in [Martinis et al., 2011] vorgeschlagen.

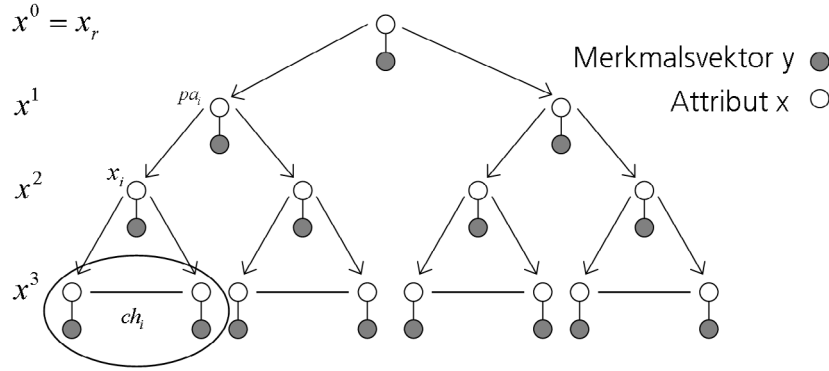


Abbildung 4.1: Dyadische Darstellung des hybriden Modells. Jeder Knoten wird mit einer Klassenzugehörigkeit x (Attribut) sowie (optional) mit einem Merkmalsvektor y assoziiert. Die Merkmalsvektoren existieren in den Ebenen des Quadrees, in denen auch Datenmodelle formuliert werden. Die Anzahl der Merkmale kann in jeder Ebene unterschiedlich sein. Die räumlich benachbarten Knoten im feinsten Level des Quadrees sind ebenfalls mit einander verbunden (Nachbarschaft erster Ordnung).

erster Ordnung folgt für die Übergangswahrscheinlichkeiten

$$P(x^n | x^{n-1}, \dots, x^0) = \prod_{i \in x^n} P(x_i | pa_i), \quad \forall n \in \{1, \dots, N\}. \quad (4.2)$$

3. Für das Beobachtungsmodell $P(y|x)$ wird eine bedingte Unabhängigkeit der Komponenten y gegeben x angenommen, so dass eine elementweise Faktorisierung der Form

$$P(y|x) = \prod_{i \in x} P(y_i | x_i) \quad (4.3)$$

vorliegt. Sollte die Beobachtung in einem Knoten oder in einem gesamten Level des Quadrees nicht existieren, so gilt $P(y_i | x_i) = 1.0$.

Aus diesen Annahmen für das *a priori*- und für das Datenmodell folgt eine faktorisierte Form der gemeinsamen Verteilung von x und y

$$P(x, y) = P(x_r) \prod_{i \neq r} P(x_i | pa_i) \prod_{i \in x} P(y_i | x_i), \quad (4.4)$$

wobei $P(x_r)$ die *a priori* Wahrscheinlichkeit der Wurzel des Quadrees, $\{P(x_i | pa_i)\}_{i \neq r}$ die Klassenübergangswahrscheinlichkeiten zwischen den Eltern- und den Kindknoten und $\{P(y_i | x_i)\}_{i \in x}$ die bedingten *likelihoods* in den verschiedenen Hierarchieebenen sind.

Für das ungerichtete planare Modell wird ein auf regulären Gittern basiertes MRF verwendet. Hier wird die übliche Positivität sowie die Markov-Eigenschaft angenommen

$$P(x) > 0, \quad \forall x \in X \quad (\text{Positivität}) \quad (4.5)$$

$$P(x_i | x) = P(x_i | x_{\mathcal{N}_i}) \quad (\text{Markov-Eigenschaft}) \quad (4.6)$$

so dass die Energiefunktion mit

$$U(x|y) = \sum_{i \in x^N} P(y_i|x_i) + \sum_{i \in x^N} \sum_{i' \in \mathcal{N}_i} \beta \delta_k(x_i, x_{i'}) \quad (4.7)$$

gegeben ist (vergleiche Abschnitt 3.1). Es folgt eine detaillierte Darstellung des Datenmodells, des *a priori* Modells sowie der Methoden zur Parameterschätzung und Inferenz für das hybride Modell.

4.2 Datenmodell

In dieser Arbeit werden sowohl für das hierarchische als auch das planare MRF GMMs für die Modellierung von $p(y_i|x_i)$ verwendet. Diese Herangehensweise ist bei generativen Ansätzen weit verbreitet und gebräuchlich. Für das planare MRF wird lediglich das Datenmodell in der feinsten Ebene des Quadrees herangezogen.

Im Gegensatz zu den meisten in der Literatur vorgestellten Ansätzen, wird in dieser Arbeit die Einführung eines Merkmalsraums für jede Ebene des Quadrees vorgeschlagen. Dies ermöglicht eine Verwendung unterschiedlicher Bildmerkmale in den verschiedenen Skalenebenen und erlaubt darüber hinaus die Modellierung unterschiedlicher thematischer Klassen in den Ebenen. Da ursprünglich keine Information über die spektralen Eigenschaften der zu parametrisierenden Klassen vorliegt, wird die Anwendung eines teilungsbasierten EM-Algorithmus [Verweridis und Kotropoulos, 2008] für das *clustering* in den einzelnen Ebenen vorgeschlagen. Auf der Grundlage einer Analyse und Beurteilung der Verteilung der Punkte einer Komponente (eines *clusters*) erfolgt die sukzessive Aufteilung dieser, so dass die Anzahl der *cluster* vor der Schätzung nicht bekannt sein muss.

Die Identifikation der jeweils relevanten Bildmerkmale erfolgt ebenfalls gesondert in jeder Ebene (Abschnitt 6.3). Prinzipiell handelt es sich dabei um eine heuristische Methode, bei der jeweils das Schätzergebnis des *clustering* für die Bewertung verschiedener Kombinationen von Merkmalen analysiert wird.

4.3 A priori Modell

Für die Modellierung der *a priori* Wahrscheinlichkeiten des hierarchischen MRFs wird in [Bouman und Shapiro, 1994], [Laferté et al., 2000] und [Martinis et al., 2011] ein Potts-Modell verwendet. Dabei wird durch

$$\forall i \in x^n, P(x_i = k | pa_i = l) = \begin{cases} \theta_n, & \text{wenn } k = l \\ \frac{1-\theta_n}{M-1}, & \text{sonst} \end{cases} \quad (4.8)$$

mit $\theta_n > 1/K$ bei K Klassen, die Gleichheit der Klasse eines Knotens x_i und seines Elternknotens pa_i favorisiert. Dieses Modell bewirkt demnach eine Glättung im Schätzergebnis, die verstärkt wird, wenn θ größer wird. Zur Vereinfachung wird oft für jede Ebene n sowie für jede Klasse k derselbe Parameter θ verwendet und manuell definiert. Bei flächendeckenden und homogenen thematischen Klassen kann dies zu guten Ergebnissen führen. Bei relativ kleinflächig auftretenden thematischen Klassen, wie beispielsweise bei schmalen Straßen oder kleinen Häusern, ist dieser Ansatz nicht zielführend, da die kleinen Strukturen durch die Glättung trotz korrektem Datenmodell verschwinden können. Die manuelle Wahl der Übergangswahrscheinlichkeiten ist nach eigener Erfahrung sehr schwierig und wird durch die Freiheit der Definition

von verschiedenen thematischen Klassen in den Hierarchieebenen zusätzlich enorm erschwert. Aus diesen Gründen soll in dieser Arbeit eine Schätzung der Übergangswahrscheinlichkeiten mittels eines EM-Algorithmus [Feng et al., 2002] erfolgen. Grundlage für die Anwendung des Schätzverfahrens ist eine ausreichende Menge an Trainingsdaten in Form von klassifizierten Bildsegmenten. Während in den bisher erschienenen Veröffentlichungen dafür meist sogar mehrere manuell klassifizierte Bilder zur Verfügung stehen, kann in dieser Arbeit lediglich auf die vom Bildinterpreten manuell angegebenen Trainingsdaten für das Datenmodell zurückgegriffen werden. Es ist dabei zu erwarten, dass diese Menge an Trainingsdaten nicht ausreichend und repräsentativ für die Schätzung der Übergangswahrscheinlichkeiten ist. Die Angabe von Trainingsdaten wird sich auf kleine repräsentative Bildbereiche in den unterschiedlichen Ebenen beschränken, so dass bestimmte Klassenkombinationen mitunter überhaupt nicht in den Trainingsdaten auftreten könnten. Daher wird in dieser Arbeit das Training auf der Grundlage eines Zwischenergebnisses vorgeschlagen. Mit Hilfe des Datenmodells kann eine lokale ML-Klassifikation durchgeführt werden, wobei aufgrund von Mehrdeutigkeiten im jeweiligen Merkmalsraum Fehlklassifikationen zu erwarten sind. Jedes Bildelement wird hier, unabhängig davon, wie gering die jeweilige maximale Wahrscheinlichkeit dafür ist, einer Klasse zugewiesen. Gründe für geringe Wahrscheinlichkeiten für alle Klassen können nicht ausreichend vorhandene Trainingsdaten sowie die Vernachlässigung von Klassen bei der Modellierung sein. Bei einer pixelbasierten ML-Klassifikation ist daher das Setzen eines Schwellwertes üblich, um die Pixel zu identifizieren, die für alle Klassen nur geringe *likelihoods* $\log p(y_i|x_i)$ aufweisen. Da auf diese Weise die Anzahl der vermutlich falsch und unsicher klassifizierten Elemente verringert werden kann, wird hier ebenfalls eine ML-Klassifikation der Bildsegmente unter Verwendung dieses Tests vorgeschlagen. Die Nullhypothese des Tests lautet

$$\begin{aligned} H_0 : \chi_{f,\alpha}^2 &\geq (y_i - \mu_k)^T \Sigma_k^{-1} (y_i - \mu_k) \\ p(y_i|x_i = k) &> p(y_i|x_i = l), \forall k \neq l \end{aligned} \tag{4.9}$$

wobei der Freiheitsgrad f der Anzahl der Dimensionen des Merkmalsraums entspricht, α das Fehlniveau und die rechte Seite der Nullhypothese (4.9) die Mahalanobis-Distanz zu einem Clusterzentrum ist. Des Weiteren wird die Nullhypothese für die Bildelemente verworfen, die nicht eindeutig zu einer Komponente zugeordnet werden können. Für Segmente, die daraufhin keine Klassenzugehörigkeit aufweisen, wird bei der EM-Schätzung eine Gleichverteilung angenommen.

Für das planare MRF wird der in Abschnitt 3.1 beschriebene ICM-Algorithmus angewendet. Da es sich um ein isotropes und homogenes Modell handelt, wird hier lediglich ein Parameter β für den Einfluss des Kontexts angesetzt. Eine Schätzung des Parameters auf der Grundlage von Trainingsdaten ist grundsätzlich möglich, wurde allerdings in dieser Arbeit nicht umgesetzt. Der Grund dafür ist die Struktur der durch obigen Test bereitgestellten Trainingsdaten. Ein Ablehnen der Nullhypothese (4.9) ist in Bereichen mit Klassenübergängen und Mischklassen wahrscheinlicher, als in homogenen Bildbereichen. Somit treten in den daraus resultierenden Trainingsdaten mitunter verschiedene Nachbarschaften von thematischen Klassen nicht oder nur im geringen Maße auf (Abbildung 4.2). Die Trainingsdaten repräsentieren demnach nicht zwingend die tatsächlich vorherrschenden Nachbarschaftsbeziehungen, woraufhin der Parameter β manuell gewählt wird.

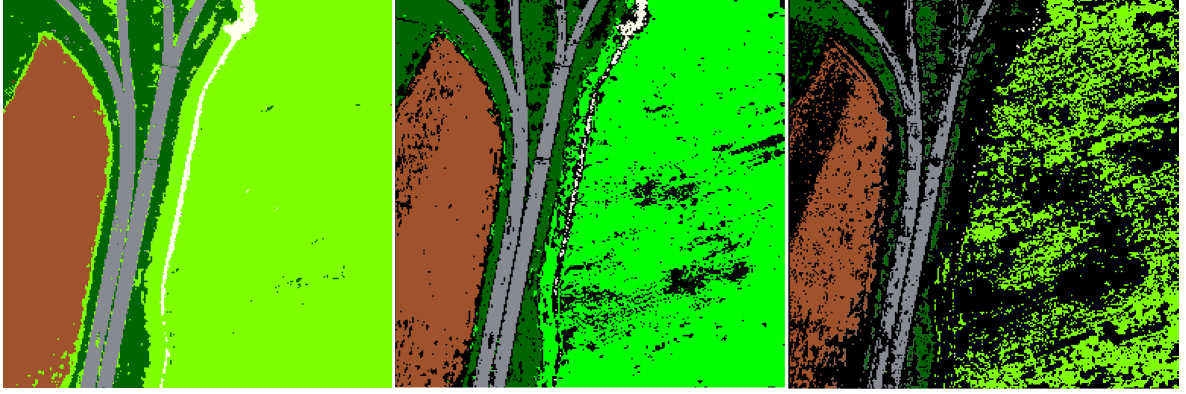


Abbildung 4.2: Links: Ergebnis einer ML-Klassifikation. Mitte: Ergebnis mit $\chi^2_{\alpha,f}$ -Test mit $\alpha = 0.1$. Rechts: Ergebnis mit $\chi^2_{\alpha,f}$ -Test mit $\alpha = 0.3$. Ursprüngliches Bild und Referenz: Abbildung 5.2.

4.4 Parameterschätzung

Die Schätzung der Parameter θ_d , θ_p für das Daten- und das *a priori* Modell bei gegebenen Trainingsdaten x_T (Klassenlabels) und y (Bildaten) erfolgt mittels ML-Training durch die Maximierung

$$\begin{aligned}\hat{\theta}^{ML} &= \arg \max_{\theta_d, \theta_p} P(x_T, y | \theta_d, \theta_p) \\ &= \arg \max_{\theta_d, \theta_p} P(y | x_T, \theta_d) P(x_T | \theta_p).\end{aligned}\tag{4.10}$$

Daran wird deutlich, dass die Parameter von Daten- und *a priori* Modell separat geschätzt werden können.

4.4.1 Parameterschätzung für das Datenmodell

Die Parameter des Datenmodells sind mit den Mittelwerten μ , den Kovarianzmatrizen Σ sowie den Gewichten π der Komponenten gegeben. Die Schätzung dieser erfolgt auf der Grundlage manuell angegebener Trainingsdaten, wobei jede Klasse und jede Hierarchieebene gesondert behandelt wird. Für die Bewertung einer geschätzten Konfiguration werden wiederum alle Komponenten in einem Merkmalsraum betrachtet. Um die Wahl der Anzahl der Komponenten zu umgehen, wird in dieser Arbeit ein teilungsbasierter EM-Algorithmus [Ververidis und Kotropoulos, 2008] angewendet. Hier wird jede Klasse mit einem *cluster* initialisiert und dieses sukzessive aufgeteilt, bis jedes *cluster* das Resultat einer einzelnen multivariaten Gauß-Verteilung ist. Die Gauß-Funktion für eine Komponente q ist

$$p(y | \mu_q, \Sigma_q) = (2\pi_q)^{-D/2} (\det \Sigma_q)^{-1/2} \exp \left\{ -\frac{1}{2} (y - \mu_q)^T \Sigma_q^{-1} (y - \mu_q) \right\}, \tag{4.11}$$

wobei D die Dimension des Merkmalsraums ist. Die Wahrscheinlichkeitsdichtefunktion von y ist mit dem GMM

$$p(y_i | \mu, \Sigma) = \sum_{q=1}^Q \pi_q p(y_i | \mu_q, \Sigma_q) \tag{4.12}$$

gegeben, wobei Q die Anzahl aller Komponenten ist. Die Initialisierung der Komponenten erfolgt durch die Berechnung von Mittelwert und Standardabweichung für jede Klasse unter Einbezug aller Trainingsdaten. Die Parameter π_q , μ_q und Σ_q werden prinzipiell für jede Klasse iterativ durch den bekannten EM-Algorithmus [Dempster et al., 1977] geschätzt:

- **E-Schritt:** Berechnung der Wahrscheinlichkeit der Zugehörigkeit zu einem Cluster q für alle Vektoren y_i in der r -ten Iteration durch

$$p_q^r(y_i) = \frac{\pi_q^{r-1} p(y_i | \mu_q^{r-1}, \Sigma_q^{r-1})}{\sum_{q'=1}^Q \pi_{q'}^{r-1} p(y_i | \mu_{q'}^{r-1}, \Sigma_{q'}^{r-1})}, \quad (4.13)$$

mit $r \geq 2$.

- **M-Schritt:** Unter Verwendung der im E-Schritt berechneten Wahrscheinlichkeiten werden die gesuchten Parameter neu berechnet.

$$\pi_q^r = \frac{1}{N} \sum_{i=1}^N p_q^r(y_i) \quad (4.14)$$

$$\mu_q^r = \frac{\sum_{i=1}^N p_q^r(y_i) y_i}{\sum_{j=1}^N p_q^r(y_j)} \quad (4.15)$$

$$\Sigma_q^r = \frac{\sum_{i=1}^N p_q^r(y_i) (y_i - \mu_q^r) (y_i - \mu_q^r)^T}{\sum_{j=1}^N p_q^r(y_j)} \quad (4.16)$$

Die beiden Schritte werden alternierend durchgeführt, bis der Erwartungswert der *log-likelihood* Funktion des Modells ein lokales Maximum erreicht. Nach Erreichen des Maximums wird überprüft, welche Komponente im nächsten Schritt aufgeteilt wird. Die Identifikation einer zu teilenden Komponente erfolgt durch einen multivariaten Test, basierend auf der Mahalanobis-Distanz eines jeden Merkmalsvektors zum Zentrum dessen Komponente. Sofern die multivariate *kurtosis* („Spitzigkeit“) der Verteilung der Mahalanobis-Distanzen der zu teilenden Komponente signifikant groß ist, werden die Zentren der beiden neuen Komponenten gleich dem alten Zentrum gesetzt, jedoch verschiedene initiale Kovarianzmatrizen gewählt. Ansonsten werden die neuen Zentren auf der Hauptachse der Ellipse der alten Komponente jeweils im Abstand der halben Länge der Hauptachse vom alten Zentrum gewählt. Im ersten Fall werden die Kovarianzmatrizen durch zufällig generierte Diagonalmatrizen, im zweiten Fall mit der Kovarianzmatrix der geteilten Komponente initialisiert. Der schematische Ablauf des Verfahrens ist in Abbildung 4.3 dargestellt.

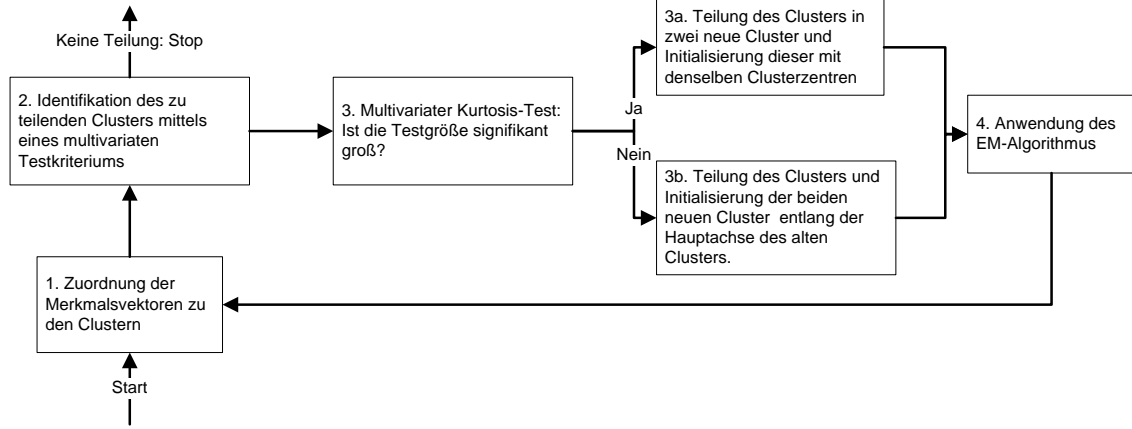


Abbildung 4.3: Schematischer Ablauf des *clusterings* nach [Ververidis und Kotropoulos, 2008].

Es werden die folgenden Schritte durchgeführt:

1. Jeder Merkmalsvektor wird entsprechend der jeweiligen Konfiguration eindeutig zu einem *cluster* zugeordnet.
2. Die Identifikation einer einzelnen zu teilenden Komponente erfolgt auf der Basis eines multivariaten Tests unter Verwendung der Information der Verteilung der Mahalanobis-Distanzen zum jeweiligen Zentrum. Die Menge der Merkmalsvektoren eines *clusters* ist $y = [y_1, \dots, y_N]$. Die Mahalanobis-Distanz eines Vektors y_i vom Zentrum μ ist definiert als

$$r_i = (y_i - \mu)^T \Sigma^{-1} (y_i - \mu). \quad (4.17)$$

Die empirische kumulative Verteilungsfunktion der Zufallsvariablen R wird $\hat{F}_R(r_i)$ bezeichnet und durch Sortieren der Mahalanobis-Distanzen in aufsteigender Reihenfolge sowie durch Setzen von $\hat{F}_R(r_i) = i/N$ definiert. Die theoretische kumulative Verteilungsfunktion $F_R(r_i)$ ist gegeben durch

$$F_R(r_i) = I_{\frac{Nr_i}{(N-1)^2}} \left(\frac{D}{2}, \frac{N-D-1}{2} \right), \quad (4.18)$$

wobei $I_x(a, b)$ die unvollständige Beta-Funktion ist. In [Ververidis und Kotropoulos, 2008] wird anhand der theoretischen kumulativen Verteilungsfunktion $F_R(r_i)$ für jeden Vektor y_i ein Konfidenzintervall $(k_{i,\lambda}^{low}, k_{i,\lambda}^{high})$ bestimmt, wobei $\lambda \in \{0.90, 0.95, 0.99\}$ gewählt wird. Die empirische und theoretische Verteilung sowie die Konfidenzintervalle sind in Abbildung 4.4 beispielhaft für ein *cluster* veranschaulicht. Hier ist die empirische kumulative Verteilungsfunktion von r_i dargestellt und den dazu korrespondierenden Konfidenzintervallen gegenüber gestellt. Für $r < 1.2$ ist die Verteilung $\hat{F}_R(r_i)$ signifikant kleiner als die theoretische Verteilung $F_R(r_i)$. Das multivariate Kriterium zählt die Anzahl d , für die $\hat{F}_R(r_i)$ außerhalb des Konfidenzintervalls $(k_{i,\lambda}^{low}/N, k_{i,\lambda}^{high}/N)$ liegt.

3. Für das zuvor identifizierte *cluster* wird nun ein multivariater *kurtosis* Test durchgeführt, um die Art der Initialisierung der beiden neuen *cluster* zu bestimmen. Die *kurtosis* ist

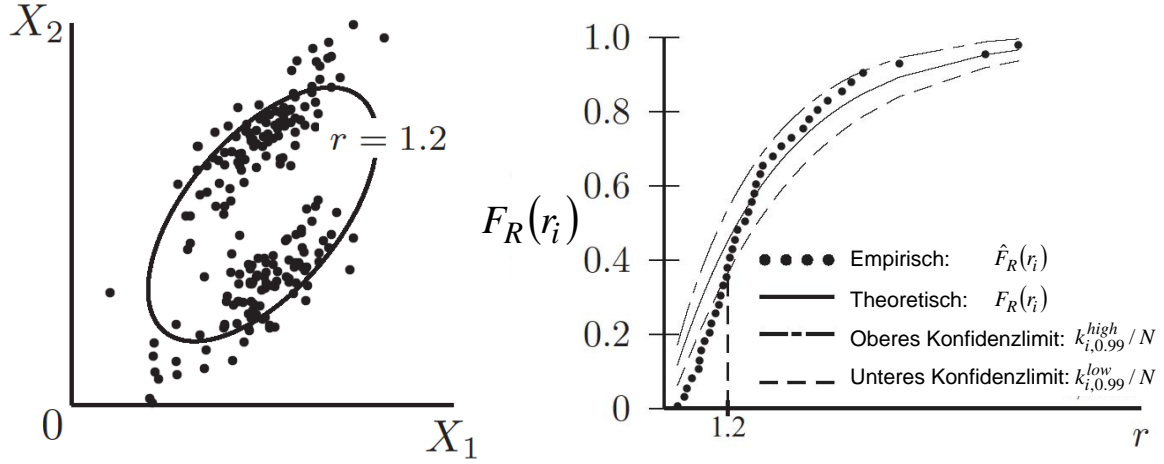


Abbildung 4.4: Links: Eine Menge von zweidimensionalen Merkmalsvektoren. Rechts: Das dazu korrespondierende multivariate Testkriterium. Quelle: [Ververidis und Koutropoulos, 2008].

nach [Marida et al., 1979] definiert durch

$$K(y) = \frac{1}{N} \sum_{i=1}^N r_i^2 \quad (4.19)$$

und stellt ein Maß für die „Spitzigkeit“ der Verteilung der Mahalanobis-Distanzen in einem Cluster dar. Dieses Maß erlaubt eine Aussage darüber, ob die Quelle eines *clusters* zwei oder mehrere Gauß-Funktionen mit dem gleichen Zentrum darstellt. Für den Test wird angenommen, dass eine multivariate Gauß-Dichte eine zu erwartende *kurtosis* K_0 hat, wobei $[K_{0;0.025}, K_{0;0.975}]$ das entsprechende Konfidenzintervall bei 95% Signifikanz darstellt. K_0 ist definiert durch

$$K_0 = E(K) = \left(1 - \frac{1}{N}\right)^2 \frac{N-1}{N+1} D(D+2). \quad (4.20)$$

Per Definition ist die Ordnung der obigen Werte

$$0 < K_{0;0.025} < K_0 < K_{0;0.975} < \infty. \quad (4.21)$$

Durch das Kriterium in Schritt 2 wird festgestellt, ob $K \in [K_{0;0.025}, K_{0;0.975}]$ gilt. Sofern diese Hypothese verworfen wird, bestehen für die Teilung die folgenden zwei Möglichkeiten.

- a) Sofern $K > K_0$ gilt, wird das entsprechende *cluster* in zwei neue *cluster* aufgeteilt und mit den selben Zentren jedoch zufällig gewählten Kovarianzmatrizen initialisiert.
- b) Sofern $K < K_0$ gilt, wird das entsprechende *cluster* in zwei neue *cluster* aufgeteilt, wobei die neuen Zentren auf der Hauptachse der Ellipse des ursprünglichen *clusters* liegen. Der Abstand der Zentren vom ursprünglichen Zentrum entspricht der halben Länge der Hauptachse des alten *clusters*.

4. Nach der Initialisierung der beiden neuen Komponenten wird der iterative EM-Algorithmus (Gleichungen (4.13)-(4.16)) angewendet.

Die Schritte 1-4 werden so lange durchlaufen, bis nach Schritt 2 kein *cluster* mehr zu teilen ist.

4.4.2 Parameterschätzung für das a priori Modell

Die Parameter θ_p (im Folgenden θ) des *a priori* Modells sind Übergangswahrscheinlichkeiten zwischen den thematischen Klassen in benachbarten Hierarchieebenen und werden mit bedingten Wahrscheinlichkeitstabellen, also in Form von Matrizen, dargestellt. Es seien $k = 1, \dots, K$ die möglichen diskreten Werte (Klassen) eines Knotens x_i und $l = 1, \dots, L$ die möglichen Klassen des Elternknotens pa_i von x_i . Der Parameter θ_{ikl} stellt einen Wert $P(x_i = k | pa_i = l)$ einer Wahrscheinlichkeitstabelle dar, wobei $\sum_l \theta_{ikl} = 1$ gilt. Äquivalent zu den MSRFs (Abschnitt 3.4) werden bei der EM-Schätzung in jeder Iteration *bottom-up* und *top-down* Nachrichten zwischen den Knoten des Baumes ausgetauscht, um die *a posteriori* Wahrscheinlichkeiten der versteckten Knoten im E-Schritt zu berechnen. Anschließend werden die erwarteten Klassenübergänge herangezogen, um im M-Schritt die Übergangswahrscheinlichkeiten neu zu berechnen [Ronen et al., 1995]. Die Neuberechnung der Übergangswahrscheinlichkeiten in einer Iteration basiert auf der Maximierung der *Baum auxiliary function* über die neu geschätzten Parameter θ'

$$Q(\theta', \theta) = \sum_{x_h} P(x_h | x_T, \theta) \log [P(x_h, x_T | \theta')], \quad (4.22)$$

wobei x_h die versteckten Variablen und x_T die Trainingsdaten sind. Aufgrund der begrenzten Verfügbarkeit von Trainingsdaten ist die gesonderte Schätzung von θ für jede Kante des Baumes nicht praktikabel, so dass für jedes Level n dieselben Parameter angenommen werden. Daraus ergibt sich der folgende Ausdruck für die Aktualisierung der Parameter

$$\theta'_{nkl} = \frac{\sum_{x_i \in x^n} P(x_{ik}, pa_{il} | x_T, \theta)}{\sum_{x_i \in x^n} \sum_{k'} P(x_{ik'}, pa_{il} | x_T, \theta)}. \quad (4.23)$$

Die gemeinsame Wahrscheinlichkeit $P(x_{ik}, pa_{il} | x_T, \theta)$ wird wiederum lokal in jedem Knoten durch das Weitergeben von Nachrichten nach [Pearl, 1988] berechnet

$$P(x_{ik}, pa_{il} | x_T, \theta) = \frac{1}{\sum_{l'} \pi(pa_{il'}) \lambda(pa_{il'})} \lambda(x_{ik}) \theta_{ikl} \pi(pa_{il}) \prod_{ch \in s_{x_i}} \lambda_{ch}(pa_{il}), \quad (4.24)$$

wobei s_{x_i} die Menge der Geschwisterknoten von x_i im selben Level und ch die Bezeichnung für einen Kindknoten ist. Die Funktion $\lambda(x)$ ist rekursiv definiert durch

$$\begin{aligned} \lambda(x) &= \prod_{i=1}^m \lambda_{ch_i}(x) \\ \lambda_{ch_i}(x) &= \sum_{l=1}^L P(ch_{il} | x) \lambda(ch_{il}), \end{aligned} \quad (4.25)$$

wobei m die Anzahl der Kindknoten von x ist. $\lambda_{ch}(x)$ wird als λ -Nachricht bezeichnet, welche von den Kindknoten ch zu seinen Elternknoten x gesendet wird. $\pi(x)$ ist rekursiv definiert

durch

$$\begin{aligned}\pi(x) &= \sum_{l=1}^L P(x|pa_l) \pi_x(pa_l) \\ \pi_x(pa) &= \beta \pi(pa) \prod_{pa' \in s_x} \lambda_{pa'}(pa),\end{aligned}\tag{4.26}$$

wobei pa_l den l -ten Zustand von Elternknoten pa und s_x die Geschwister von x sind. Die Normalisierungskonstante β garantiert, dass sich die Werte $\pi_x(pa)$ zu Eins summieren. $\pi_x(pa)$ wird als π -Nachricht, die von Knoten x zu seinem Elternknoten pa gesendet wird, bezeichnet.

Der Prozess erfordert eine initiale Definition von Zuständen in den Blattknoten sowie in der Wurzel des Baumes. Der π -Vektor in der Wurzel wird gleich der *a priori* Wahrscheinlichkeit der Klassen gesetzt. Sofern keine weitere Information zur Verfügung steht, wird hier üblicherweise eine Gleichverteilung angenommen. Des Weiteren sind die λ -Vektoren der Blattknoten (sowie auch der inneren Knoten) mit Nullen, und lediglich an der Stelle der entsprechenden vorliegenden Klasse in den Trainingsdaten, mit einer Eins besetzt. Bei unvollständigen Trainingsdaten wird jedes Element des λ -Vektors gleich $1/K$ gesetzt. Um den Wertebereich der Rechengenauigkeit nicht zu unterschreiten, werden die Vektoren $\lambda(x)$ nach ihrer Berechnung mit dem Faktor $1/\sum_x \lambda(x)$ normiert.

4.5 Inferenz

Die Inferenz der Klassenlabels erfolgt in Bayes'schen Ansätzen durch die Minimierung des Erwartungswertes einer Kostenfunktion C gegeben den Bilddaten

$$\hat{x} = \arg \min_{x'} E(C(x, x') | Y = y),\tag{4.27}$$

wobei die Kostenfunktion die Diskrepanz zwischen den jeweils geschätzten Klassenlabels x' und den „idealen“ Labels x bestraft. Basierend auf [Laferté et al., 2000] werden in dieser Arbeit für das hierarchische Markov-Modell zwei verschiedene Schätzverfahren angewendet, welche jeweils aus einer korrespondierenden Kostenfunktion resultieren. Dabei handelt es sich um die bekannten Schätzer *maximum a posteriori* (MAP) sowie *marginal posterior mode* (MPM). Die entsprechenden Schätzalgorithmen sind nicht iterativ und erfordern lediglich zwei Durchläufe durch die Quadtree-Struktur.

Basierend auf den Ergebnissen der hierarchischen Inferenz erfolgt die Inferenz für das planare MRF mittels ICM-Algorithmus (Abschnitt 3.1.2).

4.5.1 Hierarchische MAP-Inferenz

Die Kostenfunktion des MAP-Kriteriums ist definiert durch

$$C(x, x') = 1 - \delta(x, x'),\tag{4.28}$$

mit der Kronecker-Delta-Funktion δ . Der dazu korrespondierende Bayes'sche Schätzer ist mit

$$\hat{x} = \arg \max_x P(x|y) = \arg \max_x P(x, y)\tag{4.29}$$

gegeben. Für diese Schätzung kann ein dem Viterbi-Algorithmus [Forney, 1973] ähnliches Verfahren angewendet werden. Unter Verwendung der Bayes'schen Regel sowie der in Abschnitt 4.1 definierten statistischen Eigenschaften des hierarchischen Modells, ergibt sich der folgende Algorithmus²:

- **Bottom-up Durchlauf:** Berechnung und Speichern der Funktionen $P_i(pa_i)$ und $x_i^*(pa_i)$ für alle möglichen Werte von pa_i :

$$\text{Initialisierung von } i \in x^N \quad \left| \begin{array}{l} P_i(pa_i) = \max_{x_i} p(y_i|x_i)P(x_i|pa_i) \\ x_i^*(pa_i) = \arg \max_{x_i} p(y_i|x_i)P(x_i|pa_i) \end{array} \right.$$

$$\text{Rekursion für } i \in x^{N-1}, \dots, x^1 \quad \left| \begin{array}{l} P_i(pa_i) = \max_{x_i} p(y_i|x_i)P(x_i|pa_i) \prod_{t \in i^+} P_t(x_i) \\ x_i^*(pa_i) = \arg \max_{x_i} p(y_i|x_i)P(x_i|pa_i) \prod_{t \in i^+} P_t(x_i) \end{array} \right.$$

- **Top-down Durchlauf:** Sukzessive Bestimmung der MAP-Schätzung ausgehend von der Wurzel des Quadrtrees:

$$\begin{array}{l} \text{Initialisierung der Wurzel} \\ \text{Rekursion für } i \in x^1, \dots, x^N \end{array} \quad \left| \begin{array}{l} \hat{x}_r = \arg \max_{x_r} p(y_r|x_r)P(x_r) \prod_{i \in r^+} P_t(x_r) \\ \hat{x}_i = x_i^*(\hat{pa}_i) \end{array} \right.$$

Die Wahrscheinlichkeiten $P_i(\cdot)$ werden schnell sehr klein, so dass die übliche Rechengenauigkeit von Computern vor allem bei großen Bildern und demnach hohen Baumstrukturen nicht ausreichen ist (Zahlenbereichsunterschreitung). Aus diesem Grund werden jeweils die Logarithmen der Wahrscheinlichkeiten verwendet. Produkte von sehr kleinen Faktoren werden somit durch Summen von möglicherweise großen negativen Zahlen ersetzt.

4.5.2 Hierarchische MPM-Inferenz

Die Kostenfunktion der MAP-Schätzung bestraft Unterschiede zwischen den Konfigurationen x und x' , ohne zu beachten, wie groß diese Unterschiede sind. Die Kostenfunktion

$$C(x, x') = \sum_{i \in x} \delta(x_i, x'_i) \quad (4.30)$$

bewirkt hingegen eine Minimierung der globalen Anzahl an falsch klassifizierten Bildelementen und führt zum MPM-Schätzer. Dieser weist jedem Bildelement unter Einbezug aller jeweils einfließenden Daten y die wahrscheinlichste Klasse zu und lautet

$$\forall i, \hat{x}_i = \arg \max_{x_i} P(x_i|y). \quad (4.31)$$

Der Algorithmus für die MPM-Schätzung erfordert einen vorbereitenden *top-down* Durchlauf, um die später benötigten Wahrscheinlichkeiten $P(x_i)$ zu berechnen und kann wie folgt zusammengefasst werden:

- **Vorbereitender top-down Durchlauf:** Berechnung der *a priori* Randdichten in jedem Knoten durch

$$i \in x^1, \dots, x^N, P(x_i) = \sum_{pa_i} P(x_i|pa_i)P(pa_i) \quad (4.32)$$

²Eine ausführliche Darstellungen der Herleitung der Gebrauchsformeln erfolgt in [Laferté et al., 2000].

- **Bottom-up Durchlauf:** Bereitstellung der Verteilung für alle x_i und für alle Paare (x_i, pa_i) gegeben den Bilddaten $y_{d(i)}$ der einfließenden Knoten (einschließlich y_i):

$$\begin{aligned} \text{Initialisierung der Wurzeln} \quad & \left| \begin{aligned} P(x_i|y_i) &\propto p(y_i|x_i)P(x_i) \\ P(x_i, pa_i|y_i) &= \frac{P(x_i|pa_i)P(pa_i)p(x_i|y_i)}{P(x_i)} \end{aligned} \right. \\ \\ \text{Rekursion für } i \in x^{N-1}, \dots, x^1 \quad & \left| \begin{aligned} P(x_i|y_i) &\propto p(y_i|x_i)P(x_i) \prod_{t \in i^+} \sum_{x_t} \frac{P(x_t|y_{d(t)})P(x_t|x_i)}{P(x_t)} \\ P(x_i, pa_i|y_{d(i)}) &= \frac{P(x_i|pa_i)P(pa_i)p(x_i|y_{d(i)})}{P(x_i)} \end{aligned} \right. \end{aligned}$$

- **Top-Down Durchlauf:** Berechnung der vollständigen *a posteriori* Randdichten aus den zuvor bestimmten partiellen Randdichten:

$$\begin{aligned} \text{Initialisierung} \quad & \left| \begin{aligned} P(x_r|y) &\propto p(y_r|x_r)P(x_r) \prod_{t \in i^+} \sum_{x_t} \frac{P(x_t|y_{d(t)})P(x_t|x_r)}{P(x_t)} \\ \hat{x}_r &= \arg \min_{x_r} P(x_r|y) \end{aligned} \right. \\ \\ \text{Rekursion} \quad & \left| \begin{aligned} P(x_i|y) &= \sum_{pa_i} \frac{P(x_i, pa_i|y_{d(i)})}{\sum_{x_i} P(x_i, pa_i|y_{d(i)})} P(pa_i|y) \\ \hat{x}_i &= \arg \min_{x_i} P(x_i|y) \end{aligned} \right. \end{aligned}$$

Das Unterschreiten der Rechengenauigkeit ist auch hier wahrscheinlich. Ähnlich wie bei der MAP-Schätzung werden daher auch hier Logarithmen verwendet. Problematisch dabei ist das Auftreten von Summen, da dafür kein entsprechendes Logarithmusgesetz existiert. Das Umrechnen der zu summierenden Werte mit der Exponentialfunktion kann jedoch durch die Anwendung eines Rechentricks umgangen werden. Die Summe von zwei logarithmierten Werten a und b ergibt sich durch folgende Umformung zu

$$\begin{aligned} \sum(a, b) &= a + b = \log(\exp(a) + \exp(b)) \\ &= \log\left(\exp\left(\frac{a+b}{2}\right)\right) + \log\left(\exp\left(\frac{a-b}{2}\right) + \exp\left(-\frac{a-b}{2}\right)\right) \\ &= \frac{a+b}{2} + \log\left(2 \cosh\left(\frac{a-b}{2}\right)\right). \end{aligned} \tag{4.33}$$

4.5.3 Modifizierte hierarchische Inferenz

Die starre Struktur des Quadtree-Bildmodells einerseits, sowie die quadratischen Bildelemente andererseits, erhöhen die Wahrscheinlichkeit des Auftretens von Mischklassen innerhalb der Bildelemente. Im Zuge der hierarchischen Inferenz mit einem Datenmodell basierend auf GMMs sind für diese Bildelemente kleine und nicht signifikant voneinander unterscheidbare *likelihoods* $p(y_i|x_i)$ zu erwarten. Bei einer ML-Klassifikation ist es daher üblich, diese Elemente mit einem $\chi^2_{f,\alpha}$ -Test, mit dem Freiheitsgrad f und dem Fehlerniveau α zu identifizieren und nicht zu klassifizieren (vergleiche Abschnitt 4.4.2). In [Kersten et al., 2010] wird dieser Test erstmals bei der hierarchischen MAP-Inferenz angewendet. Wird die Nullhypothese des Tests (4.9) verworfen, wird der Wert $p(y_i|x_i)$ für alle Klassen gleich Eins gesetzt und somit lediglich die Information des *a priori* Modells für die entsprechenden lokalen Berechnungen einbezogen. Die Modifikation kann eine Verbesserung der Klassifikationsgüte gegenüber den

herkömmlichen Inferenzmethoden erwirken, führt allerdings mitunter zu einer Verstärkung der Blockstrukturen im Inferenzergebnis sowie zu einer Vernachlässigung feiner Strukturen, sofern das Fehlerniveau α zu groß gewählt wird. Durch den nachgeschalteten Einbezug von räumlichem Kontext können die Blockstrukturen wiederum vermindert und das Klassifikationsergebnis weiter verbessert werden. Kleine Strukturen werden hier, je nach Gewichtung β des Kontextterms mitunter weiter vernachlässigt (vergleiche Abbildung 4.5). Die Wahl der Parameter α und β sollte daher sehr sorgfältig erfolgen. Neben der MAP-Inferenz wurde im

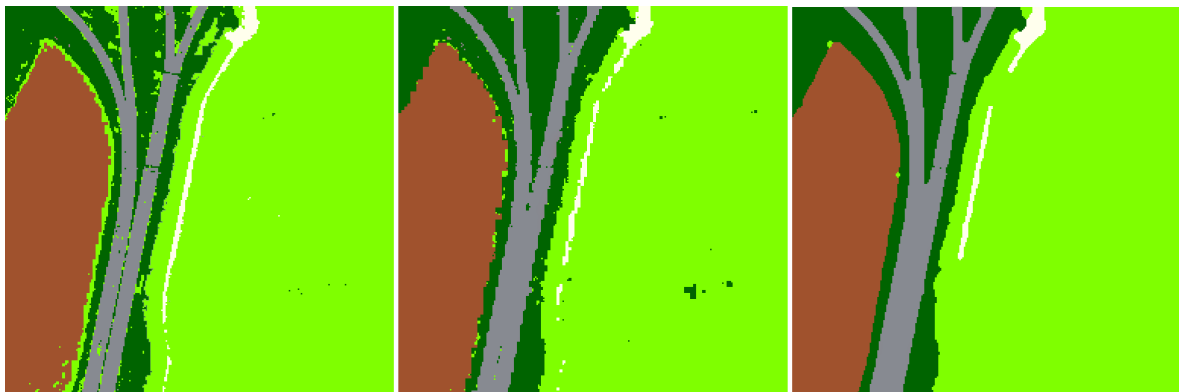


Abbildung 4.5: Ergebnisse der hierarchischen MPM-Inferenz. Links: die herkömmliche Inferenz (Gesamtklassifikationsgüte: 95.1%). Mitte: die modifizierte MPM-Inferenz (Gesamtklassifikationsgüte: 96.2%, $\alpha = 0.1$) verstärkt die Blockeffekte. Rechts: modifizierte Inferenz und anschließender ICM-Algorithmus (Gesamtklassifikationsgüte: 95.1%, $\beta = 1000$)

Rahmen dieser Arbeit auch die MPM-Inferenz durch diesen Test modifiziert. Die Wahl des Fehlerniveaus für den Test erfolgt manuell.

4.6 Bewertung des hierarchischen Inferenzergebnisses

Die Bereitstellung von Karten- und Datenprodukten im Kontext von Notfall- und Krisenkartierungen ist mit Anforderungen an die Qualität dieser verknüpft. Um den jeweiligen Nutzern der Produkte eine adäquate Entscheidungsgrundlage bereitzustellen, ist eine Bewertung der Klassifikationsergebnisse erforderlich. Aus dieser sollte ersichtlich werden, wie vertrauenswürdig das jeweilige Klassifikationsergebnis ist und in welchen Bildbereichen beziehungsweise bei welchen thematischen Klassen die Methode möglicherweise vermehrt falsche Entscheidungen getroffen hat. Als Indikator für potentielle Fehlklassifikationen wird in [Pérez et al., 2000] und [Feng et al., 2002] ein Entropiemaß der Form

$$H(x_i|y) = - \sum_{k=1}^K P(x_i|y) \log_2 [P(x_i|y)] \quad (4.34)$$

aus den bei der hierarchischen MPM-Inferenz berechneten *a posteriori* Randdichten $P(x_i|y)$ (*posterior marginals*) vorgeschlagen. Die Visualisierung der Entropien für das Inferenzergebnis in Abbildung 4.5 kann als Konfidenzkarte, in der möglicherweise falsch klassifizierte Bildelemente entsprechend eines hohen Entropiewerts hell erscheinen, aufgefasst werden. In Ab-

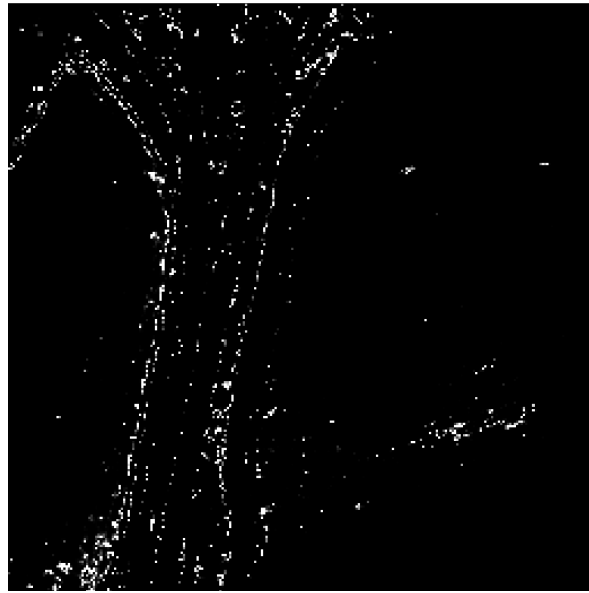


Abbildung 4.6: Entropiekarte berechnet aus den *a posteriori* Randdichten der MPM-Inferenz.

bildung 4.6 erscheinen die Bildelemente in Rand- beziehungsweise Übergangsbereichen von thematischen Klassen besonders hell. Es ist zu beachten, dass die Berechnung der Entropien bisher lediglich im Rahmen der MPM-Inferenz umgesetzt wurde, da bei der ursprünglichen MAP-Inferenz nicht alle erforderlichen Wahrscheinlichkeiten berechnet werden.

4.7 Kombination hierarchischer gerichteter und planarer ungerichteter Modelle

In der Literatur erfolgt die sequentielle Kombination von hierarchischen und planaren Modellen bisher meist durch eine starre Anwendung dieser, wobei das Ergebnis der hierarchischen Inferenz die Initialisierung für den ICM-Algorithmus darstellt. Dabei werden die beiden Modelle unabhängig voneinander betrachtet, so dass zwar die Klassenzugehörigkeiten mittels Einbezug von hierarchischem Kontext und gegebenenfalls mehrskaliger Bildinformation inferiert wird, die entsprechenden Wahrscheinlichkeiten (*a posteriori* Randdichten) jedoch beim ICM-Algorithmus unberücksichtigt bleiben. Es werden demnach aus praktischen Gründen zwei gänzlich unterschiedliche Modelle nacheinander angewendet, wobei nicht bekannt ist, welches dieser Modelle das „richtigere“ ist. Es ist also denkbar, dass ein möglicherweise gutes Ergebnis einer hierarchischen Inferenz durch den Einbezug von räumlichem Kontext wieder verschlechtert wird, oder umgekehrt. In [Martinis et al., 2011] wird erstmals eine bedingte Anwendung des ICM-Algorithmus, basierend auf den Ergebnissen der hierarchischen MAP-Inferenz vorgeschlagen, wodurch eine signifikante Reduzierung der Laufzeit erreicht wird. Hier wird eine Konfidenzkarte auf der Grundlage der Inferenzergebnisse generiert. Das Auftreten derselben Klasse in allen Knoten einer Markov-Kette des hierarchischen Graphen wird hier als Indikator für ein sicher klassifiziertes Bildsegment in der feinsten Ebene interpretiert und das entsprechende Bildsegment nicht in den ICM-Algorithmus einbezogen. Des Weiteren wird das Gewicht β des Kontextterms adaptiv auf der Grundlage der Anzahl der Unterschiede der

Klassenzugehörigkeiten zwischen dem Blattknoten und allen anderen Knoten in der entsprechenden Markov-Kette gewählt. Je weniger oft die Klasse des jeweiligen Segments bei dessen Vorfahren auftritt, desto höher ist die Wahrscheinlichkeit, dass dieses Inferenzergebnis falsch ist, woraufhin ein hoher Wert für das Gewicht β gewählt wird.

Die MPM-Inferenz weist gegenüber der MAP-Inferenz ein „besseres Verhalten“ auf [Laferté et al., 2000] und ermöglicht die Bestimmung eines Vertrauensmaßes eines jeden Knotens (vergleiche Abschnitt 4.6), der so genannten Entropie H der *a posteriori* Randdichten $p(x_i|y)$. Eigene Erfahrungen zeigen, dass sich die Ergebnisse der beiden Inferenzverfahren MAP und MPM nicht signifikant voneinander unterscheiden, so dass zu Gunsten der Bewertbarkeit der Ergebnisse durch dieses Vertrauensmaß in dieser Arbeit ausschließlich die MPM-Inferenz angewendet wird.

Die Entropie H erlaubt eine Beschränkung des ICM-Algorithmus auf die Bildelemente, die einen bestimmten Schwellwert T_H überschreiten [Martinis und Twele, 2010], wodurch signifikant verkürzte Rechenzeiten bei quantitativ sehr ähnlichen Ergebnissen erreicht werden. Im Kontext von Notfall- und Krisenkartierungen folgt daraus ein sehr attraktives Verhältnis von Kosten und Nutzen, so dass der beschränkte ICM-Algorithmus auch in dieser Arbeit angewendet werden soll. Eine visuelle Beurteilung der Ergebnisse verdeutlicht, dass offensichtliche Fehlklassifikationen, die durch den ursprünglichen ICM-Algorithmus beseitigt werden, mitunter nicht verbessert werden (vergleiche Abbildung 4.7). Andererseits kann ein besseres Verhalten bezüglich des Detailerhalts festgestellt werden. Daher wird in Abschnitt 6.9 untersucht, inwieweit die Entropie die tatsächlich falsch klassifizierten Bildelemente repräsentiert und welcher Schwellwert T_H daraufhin gesetzt werden sollte.

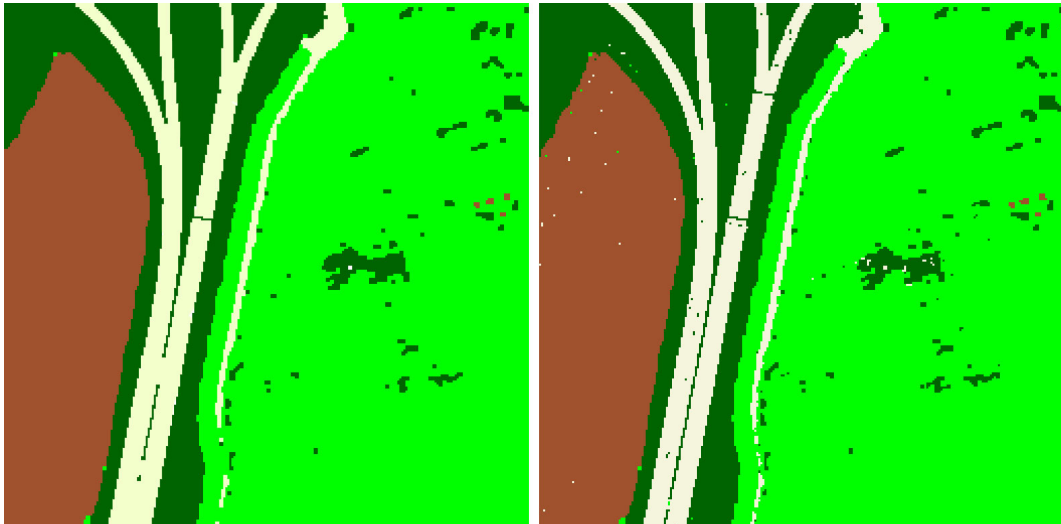


Abbildung 4.7: Ergebnisse der kombinierten hierarchischen MPM- und planaren Inferenz (ICM-Algorithmus mit $\beta = 100$). Links: Ursprünglicher ICM-Algorithmus, Gesamtklassifikationsgüte: 96.1%. Rechts: Beschränkter ICM-Algorithmus mit $T_H = 0.001$, Gesamtklassifikationsgüte: 95.9%.

Grundsätzlich sind viele verschiedene Möglichkeiten für die Integration der Information aus der hierarchischen Inferenz in den ICM-Algorithmus denkbar und wurden im Rahmen dieser Arbeit getestet. Beispielsweise könnten die *a posteriori* Randdichten als Gewicht in den ICM-Algorithmus einfließen. Damit wurden jedoch keine besseren Ergebnisse, als mit

dem ICM-Algorithmus beziehungsweise des modifizierten ICM-Algorithmus erreicht. Es ist weiterhin zu beachten, dass eine Erweiterung des Modells auch zu weiteren zu schätzenden oder manuell zu definierenden Modellparametern führt. Eine Erweiterung der Nachbarschaft durch die vier benachbarten Bildelemente in der angrenzenden Hierarchieebene darüber (Einbezug der jeweiligen Übergangswahrscheinlichkeiten) führt wiederum zu einer weiteren Verstärkung der Blockeffekte und ist daher ebenfalls nicht zielführend.

4.8 Zusammenfassung

In diesem Abschnitt wurde ein hybrides Modell der Markov-Bildmodellierung vorgestellt und die statistischen Eigenschaften sowie die jeweiligen Methoden der Parameterschätzung und Inferenz behandelt. Das hybride Modell setzt sich aus einem gerichteten hierarchischen und einem ungerichteten planaren MRF zusammen und wird aus Gründen der Effizienz sowie begründet durch die Rahmenbedingungen in dieser Arbeit sequentiell angewendet.

Die Möglichkeit der freien Definition von thematischen Klassen in den unterschiedlichen Hierarchieebenen führt dazu, dass in jeder Ebene ein separates Datenmodell (GMM) angesetzt wird. Dies erschwert eine ohnehin sehr komplexe manuelle Wahl der Parameter des *a priori* Modells, so dass diese durch entsprechende Algorithmen geschätzt werden müssen. Die dafür erforderlichen Trainingsdaten in jeder relevanten Hierarchieebene des Quadrees werden auf der Grundlage einer lokalen ML-Klassifikation bereitgestellt. Auf diese Weise wird der manuelle Aufwand für den Bildinterpretieren an dieser Stelle minimiert. Des Weiteren kann dadurch eine erste Begutachtung der Qualität der Datenmodelle durch eine visuelle Kontrolle der Zwischenergebnisse erfolgen. Um Fehlklassifikationen bei der lokalen ML-Klassifikation signifikant zu verringern, wird ein χ^2 -Test angewendet, um die unsicher klassifizierten Bildelemente zu identifizieren und bei der Schätzung der Übergangswahrscheinlichkeiten auszuschließen. Bei Ablehnung der Nullhypothese wird für das entsprechende Bildelement bei der EM-Schätzung eine Gleichverteilung für das Auftreten der thematischen Klassen angenommen.

Nach der Schätzung aller Modellparameter kann die hierarchische Inferenz erfolgen. Dafür können die beiden bekannten Schätzer MAP sowie MPM angewendet werden, wobei die *a posteriori* Wahrscheinlichkeiten der MPM-Schätzung für die Berechnung einer Konfidenzkarte verwendet werden können. Für jedes Bildelement der feinsten Ebene des Quadrees wird dafür ein Entropiemaß berechnet, welches einen Indikator für eine mögliche Fehlklassifikation darstellt. Daher wird in dieser Arbeit die ausschließliche Anwendung der hierarchischen MPM-Inferenz vorgeschlagen. Des Weiteren kann der ICM-Algorithmus auf die Bildelemente beschränkt werden, die einen bestimmten Grenzwert der Entropie überschreiten. Dies führt zu einer stark verkürzten Rechenzeit bei quantitativ ähnlichen Ergebnissen, wie mit dem ursprünglichen ICM-Algorithmus.

Aufgrund der starren Quadtree-Struktur sowie der quadratischen Bildelemente ist eine eindeutige und sichere Zuordnung der Elemente zu einer thematischen Klasse durch Mischklassen nicht immer möglich. Daher wird im Zuge einer modifizierten hierarchischen MPM-Inferenz die Regulierung des Einflusses des Datenmodells wiederum auf der Grundlage eines χ^2 -Tests vorgeschlagen. Bei Ablehnung der Nullhypothese wird für die lokalen Berechnungen der Inferenz jeweils lediglich die hierarchische Kontextinformation einbezogen und die Information des Datenmodells vernachlässigt. Auf diese Weise können Fehlklassifikationen mitunter massiv eingedämmt werden (vergleiche Abschnitt 6.8), jedoch werden die Blockstrukturen im Inferenzergebnis gegenüber der ursprünglichen MPM-Inferenz weiter verstärkt, woraufhin die anschließende Anwendung des ICM-Algorithmus in jedem Fall erfolgen sollte.

5 Integration des hybriden Modells in ein interaktives Rahmenwerk

Neben der Behandlung der Markov-Bildmodellierung ist die Konzipierung eines Rahmenwerks zur interaktiven Klassifikation unter Verwendung dieser Methodik ein weiterer Hauptbestandteil dieser Arbeit. Das Rahmenwerk wurde prototypisch in der Programmiersprache IDL (*interactive data language*) implementiert und ermöglicht eine sequentielle und interaktive Klassifikation, die Visualisierung von Zwischenergebnissen sowie den Export von Zwischen- und Endergebnissen in gängige Rasterformate. Um eine Anwendung der Methodik im Kontext von Notfall- und Krisenkartierungen zu ermöglichen, unterscheidet sich das vorgeschlagene Rahmenwerk wie folgt von anderen in der Literatur vorgestellten Methoden:

- Es können Satellitenbilder der Sensoren IKONOS, RapidEye, QuickBird, Kompsat, WorldView-2 und Spot 5 prozessiert werden. Eine Erweiterung der Methode auf weitere Bilddatenformate ist ohne großen Aufwand möglich.
- Die Größe der kleinsten (Ziel-) Bildelemente ist entsprechend der räumlichen Auflösung der Daten sowie der gewünschten thematischen Auflösung frei vom Bildinterpreten wählbar.
- Es können ad-hoc beliebige thematische Klassen in beliebigen Hierarchieebenen durch den Bildinterpreten definiert werden, wodurch die jeweilige Größe der Bildelemente, respektive die Generalisierungsstufe der jeweiligen Quadtree-Ebene berücksichtigt wird.
- Um ein hohes Maß an Flexibilität sowie eine Übertragbarkeit auf verschiedene Problemstellungen und Bildinhalte zu gewährleisten, steht eine große initiale Menge von Merkmalen für die Modellierung der Klassen in der Merkmalsraum-Domäne zur Verfügung (Datenmodell). Um redundante und irrelevante Merkmale zu identifizieren, die Dimensionalität des Klassifikationsproblems zu verringern und eine aufwändige Berechnung aller Merkmale zu vermeiden, erfolgt eine Selektion der problemrelevanten Merkmale.
- Für das Datenmodell werden GMMs verwendet. Um den jeweiligen Bildcharakteristika in den verschiedenen Skalen sowie den möglicherweise unterschiedlichen thematischen Klassen gerecht zu werden, wird je Skalenbereich ein gesondertes Datenmodell trainiert. Die Merkmalsselektion erfolgt daraufhin ebenfalls in jedem Skalenbereich gesondert.
- Die Inferenz der gesuchten Klassenlabels in der feinsten Ebene des Quadtree erfolgt mittels hierarchischer MAP- oder MPM-Schätzung. Zusätzlich dazu wird in dieser Arbeit eine modifizierte Version dieser Methoden vorgeschlagen, bei der die Fehlklassifikation von Bildelementen mit Mischklassen vermindert werden kann.
- Neben dem Einbezug von hierarchischem Kontext kann eine zusätzliche Verbesserung des Inferenzergebnisses durch den Einbezug von räumlichen Kontext in der feinsten Ebene

des Quadtrees erreicht werden. Dazu wird ein planares ungerichtetes Markov-Modell angewendet.

- Die Inferenz für das planare Modell (ICM-Algorithmus) wird auf die Bildelemente mit einer hohen Entropie der *a posteriori* Randdichten beschränkt, so dass die Laufzeit signifikant verringert wird.
- Aus den im Zuge der MPM-Inferenz berechneten Wahrscheinlichkeiten erfolgt die Berechnung einer Konfidenzkarte, die potentiell falsch klassifizierte Bildelemente ausweist.

5.1 Das Rahmenwerk

Die Datengrundlage für die Klassifikation beschränkt sich in dieser Arbeit auf das zu prozessierende Satellitenbild selbst. Für die aufgrund dessen sequentiell ablaufende Prozessierung ist ein Bildinterpret für die interaktive Steuerung der Klassifikation erforderlich. Er hat die Möglichkeit, verschiedene Prozessketten zu wählen sowie die jeweiligen Zwischenergebnisse durch visuelle Kontrolle auf Plausibilität zu überprüfen. Die sequentielle Vorgehensweise hat den Vorteil, dass etwaige Probleme, beispielsweise bei der Parameterschätzung, frühzeitig erkannt und behoben werden können. Bei einer vollautomatischen Prozesskette würde hingegen erst nach einem vollständigen Durchlauf dieser bemerkt werden, dass ein einzelner Teilprozess keine zufriedenstellenden Ergebnisse geliefert hat, wonach zunächst der entsprechende Teilprozess identifiziert werden und anschließend die Prozesskette von diesem Schritt beginnend erneut durchlaufen werden muss. Der gesamte Ablauf der Bildanalyse mit dem Rahmenwerk ist in Abbildung 5.1 dargestellt. Im ersten Schritt erfolgt die Bereitstellung der hierarchischen

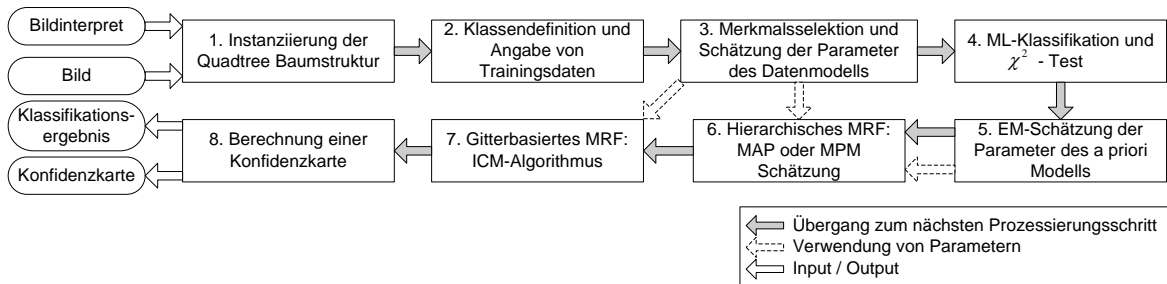


Abbildung 5.1: Schematische Darstellung der Klassifikation mit dem Rahmenwerk.

Graphenstruktur sowie der quadratischen Bildsegmente. Anschließend kann in Schritt 2 die Definition der gewünschten thematischen Klassen sowie die interaktive Angabe von Trainingsdaten für diese erfolgen. In den Schritten 3 bis 5 werden auf der Grundlage der Trainingsdaten die Parameter der verschiedenen Modelle geschätzt. Um den Aufwand für den Bildinterpret so gering wie möglich zu halten, erfolgt dies teilweise auch auf der Grundlage von Zwischenergebnissen. Anschließend erfolgt die eigentliche Klassifikation (Schritt 6 und 7), sowie die Berechnung einer Konfidenzkarte, welche ein Maß der Klassifikationsgüte darstellt (Schritt 8). Im Folgenden werden die einzelnen Schritte detailliert beschrieben.

1. Im ersten Schritt wird eine vollständige Quadtree Baumstruktur instanziiert. In Abhängigkeit der gewünschten räumlichen und thematischen Auflösung des Klassifikationsergebnisses kann der Bildinterpret die Größe des kleinsten Bildelements wählen. Die feinste

Ebene des Quadrtrees ist gleichzeitig auch die Zielebene der Klassifikation. Bei einer Anzahl von N Levels des Graphen ergeben sich (4^{N-1}) Bildelemente in der feinsten Ebene. Je kleiner die Größe gewählt wird, desto mehr Bildsegmente werden generiert, wodurch dementsprechend längere Laufzeiten bei Parameterschätzung und Inferenz zu erwarten sind.

2. Der Bildinterpret kann nun beliebige thematische Klassen in beliebigen Hierarchieebenen definieren. Dafür müssen jeweils interaktiv Trainingsdaten in Form von Bildelementen angegeben werden. Für die statistische Modellierung der Klassen im Merkmalsraum werden GMMs parametrisiert, so dass die Definition von mindestens zwei Klassen in einem Skalenbereich erforderlich ist (beispielsweise *Siedlung* und *Rest*).
3. Für jedes Bildelement wird aus den Bilddaten ein Merkmalsvektor berechnet. Um redundante und irrelevante Merkmale zu eliminieren, eine Unterrepräsentation der Klassen im Merkmalsraum aufgrund einer zu hohen Dimensionalität zu vermeiden sowie um nicht alle Merkmale für jedes Bildelement berechnen zu müssen, werden die problemrelevanten Merkmale mittels Merkmalsselektion identifiziert. Dies erfolgt durch ein heuristisches Verfahren, das die jeweils geschätzten Datenmodelle verschiedener Merkmalskombinationen miteinander vergleicht und bewertet. Die Anzahl der Dimensionen muss dabei vom Bildinterpret definiert werden.
4. Auf der Grundlage der in Schritt 3 gelernten Datenmodelle kann nun eine ML-Klassifikation in den entsprechenden Levels erfolgen. Dieser Schritt wird durchgeführt, um ausreichend Trainingsdaten für die Schätzung der Parameter des hierarchischen *a priori* Modells zur Verfügung zu stellen. Da in dem Ergebnis dieses lokalen Verfahrens mit Fehlklassifikationen zu rechnen ist, werden lediglich die Bildelemente einer thematischen Klasse zugeordnet, welche auf der Grundlage eines $\chi^2_{f,\alpha}$ -Tests eindeutig und mit einer bestimmten Sicherheitswahrscheinlichkeit $P = 1 - \alpha$ zugeordnet werden können. Das Fehlerniveau α muss wiederum vom Bildinterpret definiert werden.
5. Die lokal klassifizierten Bildelemente werden nun für die EM-Schätzung der Parameter des *a priori* Modells herangezogen.
6. Nun kann die hierarchische Inferenz erfolgen. Dafür stehen zwei gebräuchliche Methoden, die *maximum a posteriori* (MAP) sowie die *marginal posterior modes* (MPM) Inferenz, zur Verfügung. Zusätzlich dazu können in dieser Arbeit modifizierte Versionen dieser Methoden angewendet werden. Alle Inferenzalgorithmen sind exakt und nicht iterativ.
7. Um das Klassifikationsergebnis weiter zu verbessern und die bei der hierarchischen Inferenz üblicherweise auftretenden Blockstrukturen zu verringern, kann räumlicher Kontext innerhalb des feinsten Levels des Quadrtrees einbezogen werden. Dies erfolgt durch ein gitterbasiertes Markov-Zufallsfeld, wobei als initiale Konfiguration für den iterativen ICM-Algorithmus das Ergebnis der hierarchischen Schätzung dient.
8. Die aus den Wahrscheinlichkeiten der hierarchischen Inferenz abgeleiteten Entropiewerte bilden die Grundlage für die Erstellung einer Konfidenzkarte.

Für eine Prozessierung sind insbesondere die folgenden interaktiven Eingaben erforderlich:

- Schritt 1: Definition der kleinsten Größe der Bildelemente, beispielsweise 4×4 Pixel. Die kleinste mögliche Größe stellt ein Pixel selbst dar.
- Schritt 2: Definition der thematischen Klassen sowie Angabe von Trainingsdaten für jede dieser.
- Schritt 3: Vorgabe der Dimensionalität der Merkmalsräume in den verschiedenen Ebenen.
- Schritt 4: Definition des Fehlniveaus α für den χ^2 -Test sowie visuelle Überprüfung des lokalen Klassifikationsergebnisses. Ein nicht zufriedenstellendes Ergebnis ist auf eine ungünstige Anzahl an Merkmalen oder auf eine nicht repräsentative Menge an Trainingsdaten zurückzuführen. Nach etwaigen Modifikationen sind die Schritte 3 und 4 erneut durchzuführen.
- Schritt 6: Bei der Anwendung der modifizierten hierarchischen Inferenz ist wiederum die Definition des Fehlniveaus für einen χ^2 -Test erforderlich. Auch hier sollte das Ergebnis visuell überprüft werden (Blockstrukturen werden in Abhängigkeit von α verstärkt und kleine Strukturen mitunter geglättet).
- Schritt 7: Wahl des Gewichts β für den Kontextterm des planaren MRFs (ICM-Algorithmus) sowie des Schwellwertes T_H (modifizierter ICM-Algorithmus).

Es muss prinzipiell nicht die gesamte Prozesskette durchlaufen werden. Während die Schritte 1 bis 3 für alle Prozessierungsmöglichkeiten erforderlich sind, können des Weiteren die folgenden Kombinationen gewählt werden:

- Schritt 4: Lokale ML-Klassifikation aller Bildelemente.
- Schritte 4 und 7: Hier erfolgt lediglich der Einbezug von räumlichem Kontext (ICM-Algorithmus). Als initiale Konfiguration der Klassenlabels wird das Ergebnis der ML-Klassifikation verwendet.
- Schritte 4, 5 und 6: Mit dieser Kombination kann die hierarchische Inferenz mittels MAP, MPM oder einer der modifizierten Versionen dieser erfolgen.
- Schritte 4, 5, 6 und 7: Hier erfolgt die Anwendung der kompletten Prozesskette.

Die Berechnung einer Konfidenzkarte (Schritt 8) kann für alle Kombinationen, ausgenommen auf Grundlage der hierarchischen MAP-Inferenz, erfolgen.

5.2 Anwendungsbeispiel

In diesem Abschnitt erfolgt die exemplarische Anwendung des Rahmenwerks auf einen Ausschnitt der Größe 512×512 Pixel einer multispektralen IKONOS-Szene (Abbildung 5.2). Diese wurde am 6. August 2007 akquiriert und hat eine räumliche Auflösung von 1 m (näheres zu den in dieser Arbeit verwendeten Daten in Abschnitt 6.1). Es sollen beispielhaft die Klassen *Feld brach*, *Feld*, *Straße*, *Weg* und *Vegetation* klassifiziert werden. Diese Demonstration soll die



Abbildung 5.2: Links: IKONOS-Subset. Rechts: Referenz aus manueller Klassifikation.

Funktionsweise des Rahmenwerks sowie die Möglichkeiten der Prozessierung, Visualisierung und der interaktiven Steuerung aufzeigen. Prinzipiell würde in diesem einfachen Beispiel auch ein weniger komplexer Ansatz, wie beispielsweise *region growing*, zu sehr guten Ergebnissen führen. Die Nummerierung der Schritte orientiert sich an Abbildung 5.1.

1. Instanziierung der Quadtree-Baumstruktur durch Definition der Größe des kleinsten Bildelements in der untersten Stufe des Baumes (Abbildung 5.3): Hier wurde die kleinste Größe mit 2×2 Pixel definiert, so dass insgesamt $N = 9$ ($0, \dots, 8$) Levels vorliegen.

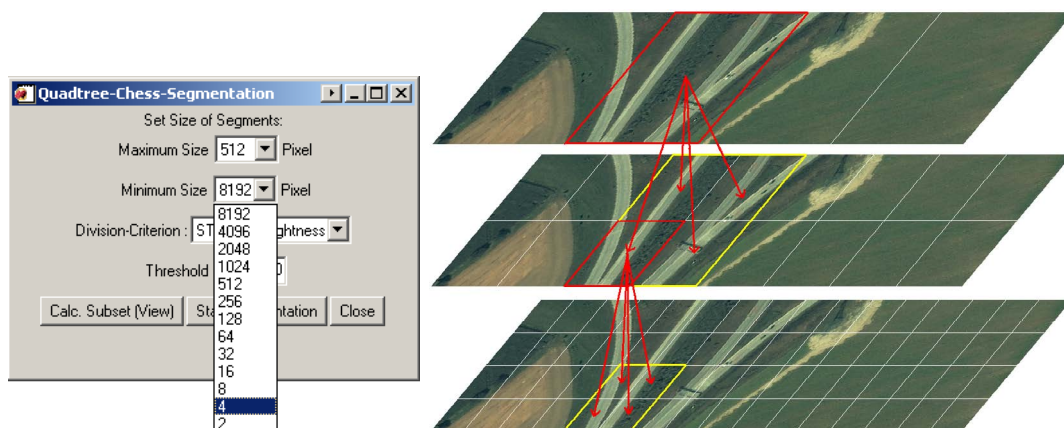


Abbildung 5.3: Links: Dialog zur Erstellung des Quadtrees. Rechts: Schematische Darstellung eines Teils der Baumstruktur.

2. Interaktive Definition der thematischen Klassen sowie Angabe von Trainingsdaten: Dies erfolgt hier in den drei feinsten Ebenen des Quadtree (Abbildung 5.4). Da in der größten der drei Ebenen die Klasse *Weg* nicht mehr durch die quadratischen Bildelemente abgebildet werden kann, wird die Klasse in dieser Ebene nicht modelliert. Die Angabe von Trainingsdaten erfolgt per Mausklick. Eine Definition durch eingrenzende Polygone ist weitaus praktikabler und sollte daher für weiterführende Untersuchungen implementiert werden.

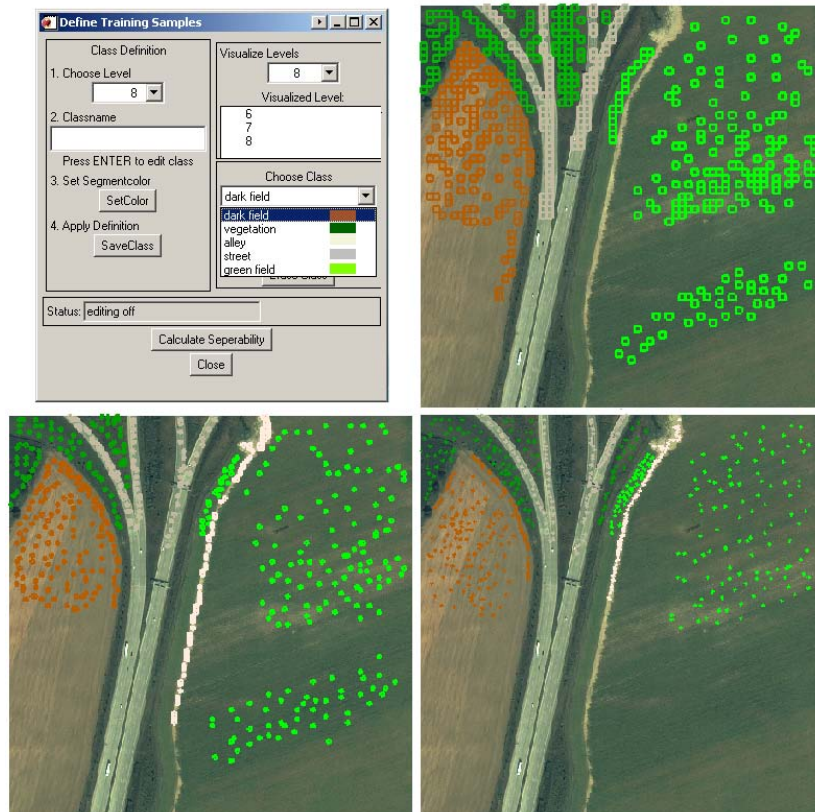


Abbildung 5.4: Oben links: Dialogfenster für die Definition und Visualisierung der thematischen Klassen. Oben Rechts: Visualisierung der Trainingsdaten in Level 6. Unten: (v. l. n. r.): Trainingsdaten in Level 7 und 8.

3. Merkmalsselektion und Schätzung der Parameter des Datenmodells: In diesem Beispiel werden jeweils sechs Merkmale in jeder Ebene gewählt. Die Merkmalsräume können anschließend durch Projektion in beliebige Ebenen visualisiert werden (Abbildung 5.5).
4. Lokale ML-Klassifikation und optionaler χ^2 -Test: In Abbildung 5.6 ist exemplarisch das Ergebnis einer ML-Klassifikation mit $\alpha = 0.5$ für alle drei Ebenen dargestellt.
5. EM-Schätzung der Parameter des *a priori* Modells (Abbildung 5.7): Die Benutzeroberfläche erlaubt die Visualisierung der Ergebnisse sowie die manuelle Veränderung der Übergangswahrscheinlichkeiten.
6. Hierarchische MAP- oder MPM-Inferenz: In Abbildung 4.5 ist ein exemplarisches Ergeb-

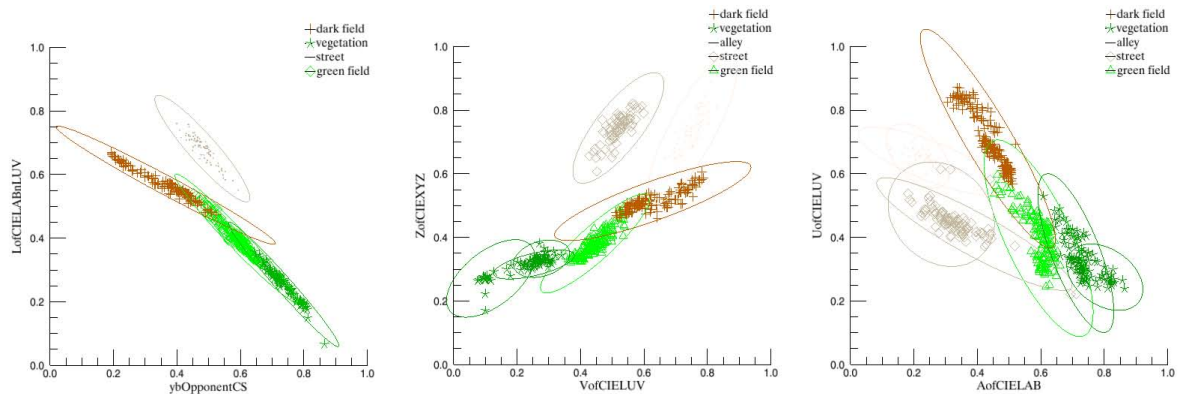


Abbildung 5.5: 2-d Visualisierung der Gauß-Modellierungen in den Merkmalsräumen der Level 6, 7 und 8 (v. l. n. r.).

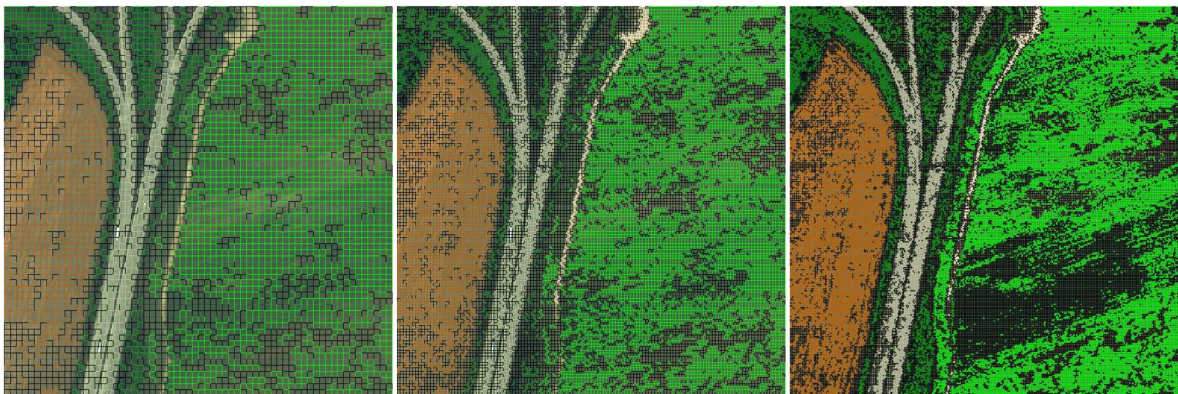


Abbildung 5.6: Ergebnis der lokalen ML-Klassifikation mit $\alpha = 0.5$ für die drei Ebenen. Bei den schwarzen Bildelementen wurde die Nullhypothese verworfen.

nis der hierarchischen MPM-Inferenz sowie der modifizierten MPM-Inferenz dargestellt (links und mitte). Es sind jeweils Blockstrukturen im Schätzergebnis zu erkennen.

7. Räumlicher Kontexteinbezug durch Anwendung des ICM-Algorithmus: Durch den Einbezug von räumlichem Kontext in der feinsten Ebene des Quadrees können sowohl Fehlklassifikationen als auch Blockartefakte deutlich verringert werden (Abbildung 4.5 rechts).
8. Bewertung der Schätzung durch Berechnung einer Konfidenzkarte (Abbildung 4.6) aus den Wahrscheinlichkeiten der hierarchischen Inferenz.

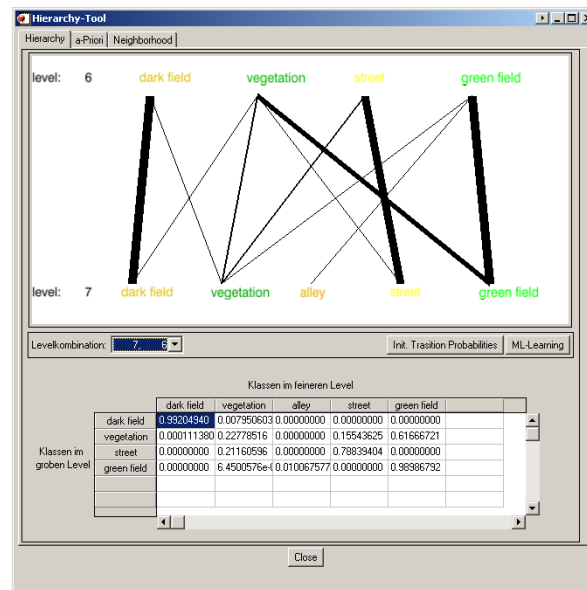


Abbildung 5.7: Visualisierung der Übergangswahrscheinlichkeiten zwischen zwei Hierarchieebenen. Je dicker die Verbindung, desto höher die entsprechende Wahrscheinlichkeit.

5.3 Zusammenfassung

In diesem Abschnitt wurde der in Kapitel 4 vorgeschlagene und beschriebene Ansatz der hybriden Markov-Bildmodellierung in ein allgemeines Gesamtkonzept zur interaktiven Klassifikation von hochauflösenden optischen Satellitenbildern integriert. Dieses Rahmenwerk besteht, aufgrund der vorliegenden Rahmenbedingungen, aus einer sequentiellen Prozesskette und ermöglicht die Klassifikation durch Kombination verschiedener Prozessschritte. Eine Klassifikation kann demnach lediglich auf der Grundlage der lokalen Bildinformation (ML-Klassifikation), unter Einbezug von hierarchischem Kontext sowie Bildinformation auf verschiedenen Skalen (hierarchische Inferenz), unter Einbezug von räumlichem Kontext (ML-Klassifikation und ICM-Algorithmus) sowie auf der Grundlage aller genannten Informationen erfolgen. Anhand eines exemplarischen Beispiels wurden die Anwendung sowie die Möglichkeiten der Visualisierung der prototypischen Implementierung veranschaulicht. Des Weiteren wurde herausgestellt, in welchen Schritten manuelle Eingaben durch den Bildinterpreten erforderlich sind.

Es ist zu erwarten, dass die Güte eines Klassifikationsergebnisses maßgeblich von den Trainingsdaten, den daraus gelernten Parametern sowie den manuell definierten Parametern abhängt. Um die Charakteristika und Anforderungen der Methode zu analysieren, folgt eine Evaluierung dieser auf der Grundlage der konkreten Anwendung auf verschiedene Satellitenbilder.

6 Evaluierung und Vergleich der Methode

Die nutzergestützte Prozessierung erfordert die Definition von thematischen Klassen, die Angabe von Trainingsdaten sowie die Definition von verschiedenen Parametern. Bezüglich der Güte der auf dieser Grundlage parametrisierten Modelle sowie des daraus resultierenden Klassifikationsergebnisses ist eine starke Abhängigkeit von diesen interaktiven Eingaben zu erwarten. In diesem Abschnitt werden daher die Eigenschaften der Methode mittels empirischer Experimente untersucht und die Robustheit des Verfahrens beziehungsweise einzelner Prozessierungsschritte bezüglich der jeweils einfließenden Parameter herausgestellt. Des Weiteren erfolgt in diesem Abschnitt der praktische Vergleich der Methode mit zwei weiteren Klassifikationsmethoden. Insbesondere werden die folgenden Punkte bearbeitet:

- Für die Schätzung der Parameter des Datenmodells ist eine ausreichende und repräsentative Menge an Trainingsdaten sowie die Definition der maximalen Anzahl d der Merkmale erforderlich. Besonders bei hochdimensionalen Merkmalsräumen ist auf eine ausreichende Menge an Trainingsdaten zu achten, da hier sehr viele Parameter geschätzt werden müssen. In Abschnitt 6.4 wird daraufhin analysiert, wie sich die Qualität von gelernten Datenmodellen in Abhängigkeit der Anzahl und Wahl der Trainingsdaten sowie der Anzahl der Merkmale verhält.
- Für die Schätzung der Parameter des *a priori* Modells werden Trainingsdaten auf der Grundlage des Datenmodells bereitgestellt und initiale Werte für die Übergangswahrscheinlichkeiten benötigt. In Abschnitt 6.5 wird untersucht, inwieweit die Schätzung der Parameter des *a priori* Modells abhängig von der Kardinalität der Trainingsdaten sowie von der Initialisierung der Übergangswahrscheinlichkeiten ist. Da die Parameterschätzung für Daten- und *a priori* Modell getrennt erfolgt, wird auch hier ein festes Datenmodell angenommen.
- Die Trainingsdaten für das *a priori* Modell werden aus dem Datenmodell generiert, wobei ein $\chi^2_{f,\alpha}$ -Test Fehlklassifikationen signifikant verringern soll. In Abschnitt 6.6 wird daher evaluiert, in welchem Maß mit diesem statistischen Test der Anteil an falsch klassifizierten Bildelementen tatsächlich verringert werden kann und wie sich diese Herangehensweise auf Parameterschätzung und Inferenz auswirkt.
- In Abschnitt 6.7 erfolgt, unter Einbezug der zuvor gewonnenen Erkenntnisse bezüglich der Parameterschätzung, eine Evaluierung der hierarchischen MPM-Inferenz. Dabei werden Datenmodelle auf verschiedenen Ebenen sowie entsprechende *a priori* Modelle trainiert und angewendet. Durch die Betrachtung und Gegenüberstellung der Ergebnisse aus verschiedenen Kombinationen von Hierarchieebenen soll der Mehrwert der Multiskalenprozessierung gegenüber einer ML-Klassifikation herausgestellt werden.
- In Abschnitt 6.8 erfolgt der Vergleich der Methode mit einer pixelbasierten ML-Klassifikation sowie mit der hierarchischen SMAP-Inferenz nach [Bouman und Shapiro, 1994].

- In Abschnitt 6.9 wird die tatsächliche Aussagekraft der Konfidenzkarte durch eine Gegenüberstellung dieser mit den tatsächlichen Fehlklassifikationen untersucht.

Für die Analysen werden insgesamt sechs verschiedene Subszenen des Sensors IKONOS untersucht, welche im folgenden Abschnitt zunächst bezüglich ihrer Charakteristika (räumliche und spektrale Auflösung, Prozessierungslevel, etc.) beschrieben werden. Bezüglich des Datenmodells ist bisher noch keine Spezifizierung der Bildmerkmale sowie der Methode der Merkmalsselektion vorgenommen worden. Da in diesem Teil der Arbeit eine konkrete Anwendung des hybriden Modells erfolgt, werden in Abschnitt 6.2 die verwendeten Merkmale und in Abschnitt 6.3 die Merkmalsselektion behandelt.

6.1 Beschreibung der Daten

Bei den Untersuchungen in diesem Abschnitt sollen die prinzipiellen Eigenschaften der einzelnen Schritte des Rahmenwerks sowie die Abhängigkeiten der Ergebnisse von den jeweils einfließenden Parametern und Trainingsdaten herausgestellt werden. Die Bilddaten weisen demnach keinen unmittelbaren Bezug zum Thema Notfall- und Krisenkartierung auf. Sie wurden vielmehr aufgrund ihrer unterschiedlichen Bildinhalte ausgewählt. Neben diesen Szenen wurden im Rahmen dieser Arbeit auch Szenen der Sensoren QuickBird [Kersten et al., 2010], RapidEye, Kompsat, WorldView-2 (simulierte Daten) und Spot-5 mit der Methode prozessiert. Die Beschaffung von Referenzdaten gestaltet sich aufgrund der räumlichen Auflösung (beispielsweise RapidEye: 5 m) sowie der jeweils aufgenommenen Region (beispielsweise Caprivi, Namibia) mitunter sehr schwierig, so dass sich an dieser Stelle lediglich die IKONOS-Szenen eignen.

Für jede der Subszenen wurden relevante thematische Klassen definiert und entsprechende Referenzklassifikationen mittels visueller Interpretation bereitgestellt. Der Prototyp der Methode erlaubt, aus Gründen einer suboptimalen Speicherverwaltung und Implementierung, eine Prozessierung von Szenen einer maximalen Größe von 1024×1024 Pixeln. In den Abbildungen 6.1 und 6.2 sind die ausgewählten Subszenen samt den Ergebnissen der visuellen Klassifikation dargestellt. Die Ergebnisse der visuellen Klassifikationen werden im Rahmen der Evaluierung als Referenz für die Bewertung der verschiedenen Ergebnisse herangezogen. Des Weiteren werden anhand dieser für die Untersuchungen bezüglich des Daten- und des *a priori* Modells zufällige Trainingsdaten generiert. Die Subszenen stammen aus insgesamt drei IKONOS-Szenen mit den folgenden Eigenschaften:

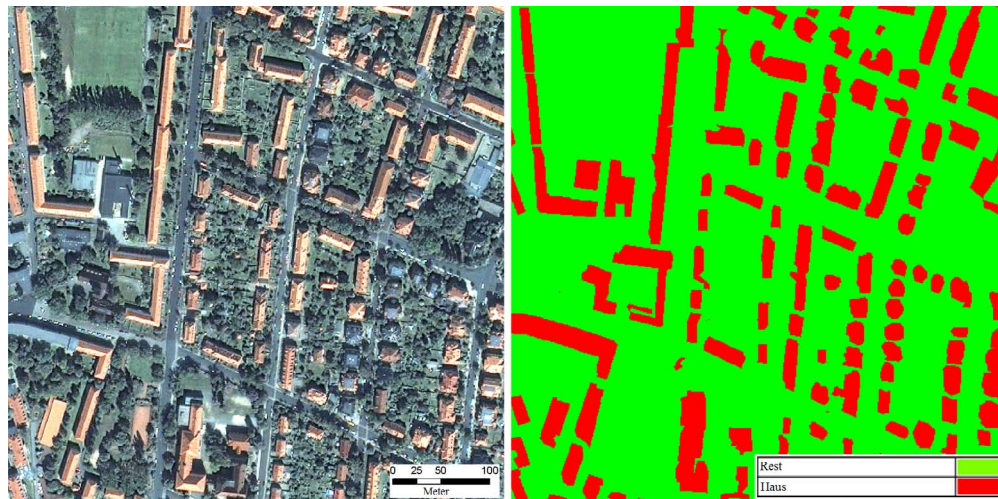
Szene	(a)	(b)	(c)	(d)	(e)	(f)
Beschreibung	Ländlich	Stadtgebiet	Stadtgebiet mit Wasser	Industrie	Ländlich mit Siedlung	Stadtgebiet mit Vegetation und Wasser
Aufnahmezeitpunkt	06.08.2007				21.08.2002	28.06.2008
Spektrale Auflösung	Multispektral, vier Kanäle: Blau, Grün, Rot und nahes Infrarot					
Geometrische Korrektur	Standardverfahren beim Prozessierungslevel <i>Geo</i> . Rektifizierung auf ein definiertes Ellipsoid und einer Kartenprojektion. Horizontale Genauigkeit: $\pm 50\text{ m}$ (CE 90 %)					
Bodenabtastdistanz PAN	0.78 m bei Querabtastung, 0.92 m bei Längsabtastung,				Quer: 1.02 m Längs: 0.98 m	Quer: 0.94 m Längs: 1.08 m
Bodenauflösung	1 m nach Pan-Sharpning (<i>cubic convolution</i>)				4 m	

Tabelle 6.1: Metadaten der prozessierten Szenen.

Alle Szenen wurden direkt mit den obigen Eigenschaften prozessiert, ohne zusätzliche radiometrische oder geometrische Korrekturen vorzunehmen. Untersuchungen bezüglich des Einflusses radiometrischer Vorverarbeitungsschritte, beispielsweise einer atmosphärischen Korrektur, auf die Güte der Klassifikationsergebnisse könnten für einzelne Anwendungsfälle durchaus relevant sein, sollen jedoch nicht Gegenstand dieser Arbeit sein.



(a) Ländliches Gebiet in der Nähe von Dresden.



(b) Stadtgebiet von Dresden.



(c) Stadtgebiet von Dresden mit Wasser.

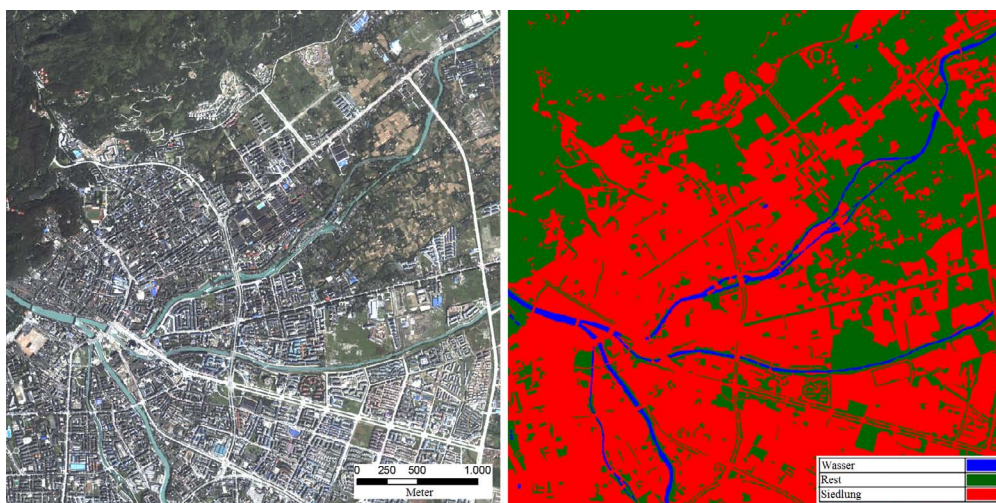
Abbildung 6.1: Subszenen der Größe 512×512 Pixel mit den entsprechenden Referenzklassifikationen aus visueller Interpretation.



(d) Ländlicher Raum mit Industrie.



(e) Ländliches Gebiet im Raum Dresden mit vereinzelt besiedelter Fläche.



(f) Stadtrand von Dujiangyan, Provinz Sichuan, China, mit Siedlungsfläche, Vegetation und Wasser.

Abbildung 6.2: Subszenen der Größe 1024×1024 Pixel mit den entsprechenden Referenzklassifikationen aus visueller Interpretation.

6.2 Bildmerkmale

Um die Anwendbarkeit der Methode auf verschiedene Problemstellungen gewährleisten zu können, wird ein umfassendes Repertoire von Bildmerkmalen für das Datenmodell zur Verfügung gestellt. Während bei den meisten Methoden in der Literatur eine feste Menge ausgewählter Merkmale verwendet wird, soll hier eine möglichst große Abdeckung des Spektrums von Merkmalen erfolgen. Das Auswahlkriterium für eine Anwendung und Implementierung ist mit der Relevanz bezüglich der Klassifikation hochauflösender Satellitenbilder gegeben. Die momentan vorliegende Implementierung des Rahmenwerks ist, dank objektorientierter Programmierung, mittels weniger Zeilen Quellcode um neue Merkmale erweiterbar. In Tabelle 6.2 wird ein Überblick über die verwendeten Merkmale gegeben. Im Anhang werden diese näher beschrieben.

Bezeichnung	Bildmerkmal	Referenz
Statistische Merkmale erster Ordnung	Mittelwert Varianz <i>skewness</i> <i>kurtosis</i> Energie Entropie	[Meterka und Strzelecki, 1998]
Ratios	NDVI und andere	-
Weitere Merkmale	Maximaler Pixelwert Minimaler Pixelwert Maximale Differenz	-
Farbraumtransformationen	HSV-Farbraum HLS-Farbraum XYZ, CIELab und CIELuv <i>Opponent</i> -Farbraum Ohta-Farbraum <i>Log-Opponent</i> -Farbraum Unbenannte Transformation Rot/Cyan-Farbigkeit P1P2-Farbraum l_1, l_2, l_3 -Farbraum $h_1 h_2 h_3$ -Farbraum	[Smith, 1978] [Luong, 1993] [Levkowitz, 1997] [Ballard und Brown, 1982] [Ohta et al., 1980] [Faugeras, 1979] [Trias-Sanz et al., 2008] [de Wouwer et al., 1999] [Vertan et al., 2000] [Gevers und Smeulders, 1999] [Wesolkowski, 1999]
Texturmerkmale	GLCM-Mittelwert GLCM-Varianz GLCM-Energie GLCM-Korrelation GLCM-Entropie GLCM-Kontrast GLCM-Homogenität GLCM-Cluster Farbtiefe GLCM-Cluster Prominenz	[Unser, 1986]
	Strukturkomplexität	[Baillard, 1997]
	<i>module</i> <i>state</i>	[Lowitz, 1983]

Tabelle 6.2: Zusammenstellung der verwendeten Bildmerkmale.

6.3 Identifikation relevanter Merkmale

Bei der Prozessierung einer IKONOS-Szene mit vier spektralen Kanälen können nach Abschnitt 6.2 insgesamt 77 Merkmale für jedes Segment berechnet werden. Die jeweils relevanten Merkmale sind in der Regel nicht bekannt, was zu den folgenden Problemen führt:

- Es ist zu erwarten, dass viele dieser Merkmalskandidaten - entweder teilweise oder komplett - irrelevant oder redundant hinsichtlich der Zielstellung sind [Dash und Liu, 1997].
- Auf der Grundlage einer Bewertung der Trennbarkeit von zwei Klassen bezüglich einzelner Merkmale (z.B. auf der Grundlage der Bhattacharyya-Distanz, wie in [Nussbaum et al., 2006]) ist kein Rückschluss auf die Unterscheidbarkeit dieser in einem mehrdimensionalen Merkmalsraum möglich. Eine Variable, welche komplett nutzlos für sich allein gesehen ist, kann einen signifikanten Genauigkeitsgewinn bewirken, wenn sie mit anderen kombiniert wird [Guyon, 2003].
- In der Regel ist die Anzahl der Trainingsdaten begrenzt, so dass durch den „Fluch der Dimensionen“ [Köppen, 2000] die Klassen bei einer hohen Dimensionalität des Merkmalsraums unterrepräsentiert sind.

Eine Reduktion der Dimensionalität des Klassifikationsproblems auf ein Minimum ist daher erforderlich, um eine optimale Klassifikationsgüte zu erreichen und den Aufwand für die Klassifikation zu verringern. Die Verfahren der Dimensionsreduktion können in die Gruppen Merkmalsselektion und Merkmalstransformation eingeteilt werden (vergleiche beispielsweise [Jain et al., 2000]).

Im Zuge der Merkmalstransformation werden durch Transformation oder Kombination der ursprünglichen Merkmale neue Merkmale bestimmt. Die Hauptkomponentenanalyse ist der wohl bekannteste Ansatz der linearen Merkmalstransformation. Grundgedanke ist dabei die Transformation von Daten in eine Basis, bei der die einzelnen Dimensionen keine Korrelation aufweisen. Der Betrag der Eigenwerte der Kovarianzmatrix gibt Auskunft über die Varianz in jeder Zieldimension. Ein optimaler Unterraum wird durch die Basisvektoren definiert, welche entlang ihrer Dimension eine große Varianz aufweisen. Die zu diesen Dimensionen zugehörigen Eigenvektoren werden anhand dieses Kriteriums zu einer Matrix zusammengestellt, welche die Transformation eines Datenpunktes vom ursprünglichen Merkmalsraum in die neue Basis definiert. Ein genereller Nachteil bei der Merkmalstransformation besteht in der Unabhängigkeit zum eigentlich verwendeten Klassifikator. Des Weiteren ist zu beachten, dass hier eine physikalische Interpretation bezüglich des Bildinhaltes und der dafür optimalen ursprünglichen Merkmale nicht möglich ist. Einer möglicherweise sogar besseren Trennbarkeit von Klassen als Resultat der Merkmalstransformation steht einem stark reduzierten Berechnungsaufwand im Zuge der Merkmalsselektion entgegen, da hier lediglich die relevanten Merkmale für alle zu klassifizierenden Elemente berechnet werden müssen [Jain et al., 2000]. Aus diesen Gründen soll in dieser Arbeit ein Verfahren aus der Gruppe Merkmalsselektion verwendet werden.

In [Dash und Liu, 1997] wird ein Überblick über die Verfahren der Merkmalsselektion zwischen 1970 und 1997 gegeben. Insgesamt werden dabei 32 Verfahren auf der Grundlage ihrer Eigenschaften bezüglich der Generierung und Evaluierung von Merkmalsuntermengen in verschiedene Kategorien eingeteilt. Eine Evaluierungsfunktion hat die Aufgabe, die Trennbarkeit verschiedener Klassen bezüglich eines oder mehrerer Kriterien zu messen. Während in [Dash und Liu, 1997] eine feinere Untergliederung der Evaluierungsfunktionen vorgenommen wird,

soll an dieser Stelle die in [Langley, 1994] und [John et al., 1994] vorgeschlagene Einteilung in die Gruppen *filter* und *wrapper* herangezogen werden. Ein Beispiel für einen *filter*-Ansatz ist die Wahl einer Merkmalskombination auf der Grundlage von Distanzmessungen im Merkmalsraum.

Für Ansätze dieser Gruppe ist anzumerken, dass die Effekte der abgeleiteten Merkmale auf den eigentlichen Klassifikationsprozess quantitativ nicht abgeschätzt werden können [John et al., 1994]. Es kann aus diesem Grund für keine der *filter*-Methoden eine Aussage über die zu erwartende Güte der Klassifikation mit den als optimal identifizierten Merkmalen getroffen werden [Dash und Liu, 1997]. Allerdings kann, beispielsweise auf der Grundlage von Distanzmaßen, eine qualitative Aussage getroffen werden. *Wrapper*-Methoden verwenden als Evaluierungsfunktion die Fehlklassifikationsrate des jeweiligen Klassifikators selbst und ermöglichen damit eine optimale Bewertung der Klassifikationsgüte. Eigene Erfahrungen mit diesem Ansatz bestätigen, dass der rechnerische Aufwand hier aufgrund der Kreuzvalidierung ungleich höher als bei den *filter*-Methoden ist. Daher wird hier trotz der suboptimalen Eigenschaften bezüglich der Evaluierungsfunktion ein *filter*-Ansatz angewendet.

Die Wahl von Merkmalsuntermengen kann prinzipiell auf drei Wegen erfolgen: komplett, heuristisch und zufällig. Bei einer vollständigen Suche werden alle $\binom{d}{n}$ möglichen Merkmalskombinationen evaluiert, wobei d die Anzahl der ursprünglichen Dimensionen und n die Anzahl der Dimensionen des reduzierten Merkmalsraums darstellt. Bei diesem Ansatz wird in jedem Fall die optimale Merkmalskombination gefunden, was allerdings selbst für moderat große Werte d und n zu rechenintensiv ist.

Mit dem *branch-and-bound*-Algorithmus [Narendra und Fukunaga, 1977] kann ein optimaler Satz von Merkmalen sehr viel schneller gefunden werden als bei einer vollständigen Suche, wobei eine monotone Selektionsfunktion angesetzt wird (das Hinzufügen neuer Merkmale darf den Wert der Evaluierungsfunktion nicht verringern). Bei einer kleinen Menge von Trainingsdaten ist dieses Kriterium aufgrund des Fluch der Dimensionen Phänomens nicht immer gewährleistet [Jain et al., 2000]. Ein Beispiel für einen zufallsbasierten Ansatz sind genetische Algorithmen [Siedlecki und Sklansky, 1989], die auf dem natürlichen Prozess der Evolution basieren. In jeder Iteration wird eine feste Anzahl möglicher Lösungen generiert, indem verschiedene genetische Operatoren im Zuge eines stochastischen Prozesses angewendet werden. Diese Operatoren bewirken beispielsweise Rekombinationen, Kreuzungen und Mutationen. Einen entscheidenden Nachteil dieses Ansatzes stellen die verschiedenen Parameter dar, für die keine festen Regeln hinsichtlich ihrer Spezifikation existieren.

Ohne auf weitere Methoden näher einzugehen, kann auf der Grundlage der vergleichenden Studien [Kudo und Sklansky, 2000], [Ferri et al., 1994] und [Jain et al., 2000] konstatiert werden, dass mit den heuristischen Verfahren SFS (*sequential forward search*) [Whitney, 1971] und SFFS (*sequential forward floating search*) [Pudil et al., 1994] nahezu optimale Lösungen bei einer tragbaren Laufzeit gefunden werden können. Daher wird die Anwendung dieser Verfahren vorgeschlagen.

SFS und SFFS-Verfahren

Das SFS-Verfahren [Whitney, 1971] beginnt mit einem leeren Satz von Merkmalen. In jedem Iterationsschritt werden alle noch nicht ausgewählten Merkmale untersucht und das Merkmal ausgewählt, welches den signifikantesten Anstieg der Evaluierungsfunktion in Kombination mit den vorher ausgewählten Merkmalen bewirkt. Auf diese Weise können allerdings lediglich suboptimale Ergebnisse erwartet werden, da einmal ausgewählte Merkmale später nicht wieder

aus dem Merkmalsatz entfernt werden können (*nesting*-Effekt).

Das SBS-Verfahren [Marill und Green, 1963] ist das *top-down* Pendant zum SFS-Ansatz: Ausgangspunkt ist hier der komplette Satz von Merkmalen, welcher in jedem Iterationsschritt reduziert wird.

Das SFFS-Verfahren [Pudil et al., 1994] wirkt dem *nesting*-Effekt entgegen, da hier sowohl das Hinzufügen, als auch das Entfernen von Merkmalen ermöglicht wird. Nachdem ein Merkmal zum aktuellen Satz hinzugefügt wurde, wird in einer *backtracking*-Phase das Entfernen einer variablen Anzahl von Merkmalen ermöglicht. Diese Abfolge wird so lange wiederholt, bis eine definierte Anzahl D an Merkmalen gefunden wurde. Beginnend mit einer Menge von k Merkmalen gestaltet sich der Ablauf in vereinfachter Weise wie folgt (Abbildung 6.3):

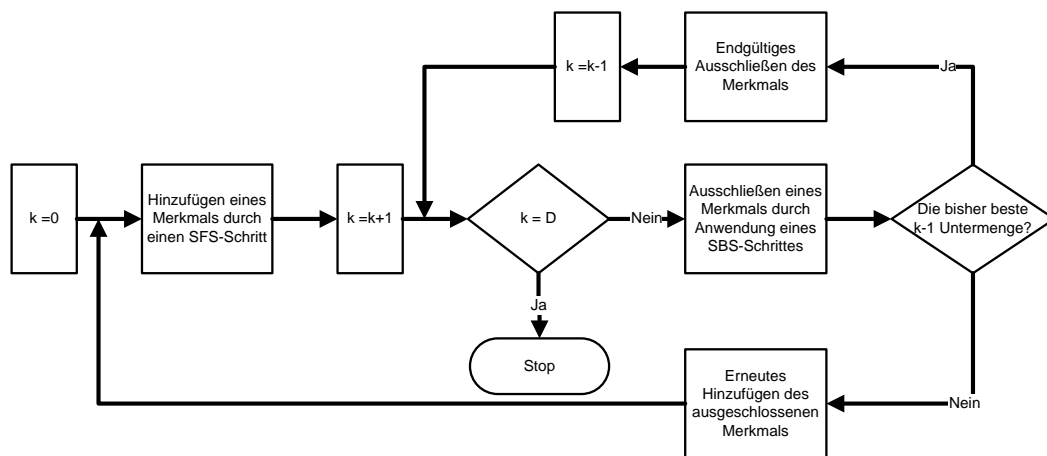


Abbildung 6.3: Ablauf des SFFS-Verfahrens nach [Somol et al., 1999].

Zunächst wird das Merkmal, welches den signifikantesten Anstieg der Evaluierungsfunktion bewirkt, durch einen Schritt des SFS-Verfahrens hinzugefügt. Es gilt nun $k = k + 1$. Das am wenigsten signifikante Merkmal des aktuellen Satzes wird nun durch die Anwendung eines SBS-Schrittes identifiziert. Ist der auf diese Weise erhaltene Satz der bisher beste mit der jeweiligen Anzahl an Merkmalen, so wird das entsprechende Merkmal entfernt und ein weiterer SBS-Schritt eingeleitet. Des Weiteren gilt $k = k - 1$. Ist der Satz mit $k - 1$ Merkmalen nicht besser als ein vorheriger, so wird das durch den SBS-Schritt identifizierte Merkmal wieder in den aktuellen Satz aufgenommen und ein neuer Iterationsschritt eingeleitet. Für die Wahl des ersten Merkmals wird in dieser Arbeit die maximale mittlere Bhattacharyya-Distanz zwischen den thematischen Klassen herangezogen. Die Iteration wird abgebrochen, sobald eine definierte Anzahl D an Merkmalen gefunden wurde.

Für die Evaluierung einer Merkmalskombination wurde im Rahmen dieser Arbeit zunächst eine Kreuzvalidierung unter Verwendung des jeweils durch den angewendeten *clustering* Algorithmus geschätzten Datenmodells verwendet. Eine großen Menge an Merkmalen und Trainingsdaten führt dabei zu sehr langen Laufzeiten, so dass verschiedene Distanzmaße für die Evaluierung als Alternative getestet wurden. Vergleiche mit den Maßen Kullback-Leibler-Divergenz, Batthacharyya-Distanz und der Mahalanobis-Distanz ergaben, dass letztere die besten Ergebnisse liefert. Daher wird die mittlere Mahalanobis-Distanz zwischen den *clustern* unterschiedlicher Klassen als Evaluierungsfunktion angesetzt. Um eine Vergleichbarkeit

der mittleren Mahalanobis-Distanzen zwischen Merkmalsräumen, die von gleich vielen jedoch unterschiedlichen Merkmalen aufgespannt werden, zu gewährleisten, werden diese mit [Aksoy und Haralick, 2000]

$$\tilde{y} = \frac{\frac{y-\mu}{3\sigma} + 1}{2}, \quad (6.1)$$

mit dem Mittelwert μ und der Standardabweichung σ des Merkmals bezüglich aller Trainingsdaten, auf das Intervall $[0, 1]$ normiert.

Die Evaluierungsmethode von Selektionsalgorithmen ist von entscheidender Bedeutung für die Bewertung dieser. In einigen Studien (beispielsweise [Kudo und Sklansky, 2000]) werden die mittels eines bestimmten Verfahrens detektierten Merkmalsmengen mit Testdaten validiert, welche auch vom Suchalgorithmus selbst verwendet wurden. Eine Gegenüberstellung des SFFS-Algorithmus und dem recht einfachen und schnelleren SFS-Verfahren [Whitney, 1971] auf Grundlage einer unabhängigen Evaluierung des gefundenen Merkmalssatzes in [Reunanen, 2003] relativiert die oft vermeintlich besseren Ergebnisse der SFFS-Methode. Bei dieser Testanordnung liefern die mittels des SFFS-Verfahrens detektierten Merkmale für sechs der sieben untersuchten Datensätze lediglich bei weniger als der Hälfte dieser bessere, und in den restlichen Fällen äquivalente Klassifikationsergebnisse. Es ist allerdings zu beachten, dass bei einem der sieben Datensätze sehr viel bessere Ergebnisse mit dem SFFS-Verfahren erzielt werden konnten, was durch die Vermeidung des *nesting*-Effekts begründet wird. Um diesen Effekt gänzlich zu vermeiden, wird in dieser Arbeit ausschließlich der SFFS-Algorithmus angewendet.

6.4 Eigenschaften der Parameterschätzung für das Datenmodell

Der Bildinterpret muss für die Klassifikation Trainingsdaten und die maximale Anzahl der Merkmalsraumdimensionen definieren. Das Ziel der Evaluierung in diesem Abschnitt ist, etwaige Zusammenhänge zwischen der Kardinalität der Trainingsdaten sowie der Dimensionalität des Merkmalsraumes einerseits und der Güte der Modellierung andererseits herauszustellen. Des Weiteren soll evaluiert werden, wie robust die Parameterschätzung unter Verwendung von unterschiedlichen Trainingsdaten ist. Zu diesem Zweck wird mit den zuvor beschriebenen IKONOS-Szenen eine Kreuzvalidierung durchgeführt, bei der jeweils Datenmodelle für die Bildelemente der Größe 2×2 Pixel geschätzt und bewertet werden sollen. Die Menge aller Bildelemente einer Szene wird dafür bei jedem Schritt der Kreuzvalidierung in die beiden Mengen Test- und Trainingsdaten aufgeteilt. Während die Trainingsdaten ausschließlich für die Parameterschätzung verwendet werden, erfolgt die Bewertung des resultierenden Datenmodells auf der Grundlage der übrigen Bildelemente (Testdaten). Die Bewertung erfolgt durch die Gesamtklassifikationsgüten, die mit dem jeweiligen erlernten Datenmodell durch eine lokale ML-Klassifikation erreicht werden.

Vor dem Hintergrund einer möglichst effizienten Prozessierung soll dieses Experiment aufzeigen, wie viele Merkmale für die unterschiedlichen Bildtypen optimale Ergebnisse liefern. Der Versuchsaufbau gestaltet sich wie folgt: Es werden jeweils unterschiedliche Mengen von Merkmalen $d_M = 2, 4, 6, 8, 10$ sowie unterschiedliche Mengen an Trainingssegmenten d_T je Klasse für das Training eines Datenmodells verwendet ($d_T = 50, 100, 150, 200, 250, 300$ für die kleinen Szenen und $d_T = 50, 100, 150, 200, 250, 300, 350, 400$ für die großen Szenen), woraus insgesamt 30 beziehungsweise 40 verschiedene Konfigurationen je Szene resultieren. Für jede dieser Konfigurationen werden im Rahmen einer Kreuzvalidierung auf der Grundlage der Referenzen 20 mal zufällige und unabhängige Trainingsdaten gewählt und mit diesen die Merkmalsselektion

sowie die Parameterschätzung durchgeführt. Daraus resultieren insgesamt 600 beziehungsweise 800 verschiedene Konfigurationen von Trainingsdaten je Szene. Als Maß für die Robustheit der Parameterschätzung wird die Standardabweichung σ der Gesamtklassifikationsgüten aus den jeweils 20 zufällig gewählten Trainingsdatensätzen der 30 beziehungsweise 40 Konfigurationen herangezogen. Neben den Gesamtklassifikationsgüten werden für die Analyse auch die Konsumenten- und Produzentengenauigkeiten der einzelnen thematischen Klassen betrachtet. Anhand dieser beiden Größen können die Ergebnisse bezüglich der einzelnen Klassen detaillierter interpretiert werden. Die Konsumentengenauigkeit (*users accuracy*, UA) drückt aus, wie wahrscheinlich ein klassifiziertes Bildsegment tatsächlich der entsprechenden Klasse angehört. Es wird dafür anhand der Referenz überprüft, wie viel Prozent der Segmente in einem Klassifikationsergebnis tatsächlich aus der Menge der jeweiligen Klasse stammen. Je mehr sich eine Klasse in einem Klassifikationsergebnis aus Referenzsegmenten anderer Klassen zusammen setzt, desto niedriger ist dieses Maß. Die Produzentengenauigkeit (*producers accuracy*, PA) bezieht sich dagegen auf die Referenzdaten beziehungsweise auf die Verteilung der Referenzsegmente durch den Klassifikator in die verschiedenen Klassen und beschreibt, wie viele der Referenzsegmente richtig klassifiziert wurden. Je mehr Segmente der Referenz falsch klassifiziert werden, desto niedriger ist dieses Maß. Es folgt eine detaillierte Darstellung ausgewählter Ergebnisse, sowie eine Zusammenfassung der Ergebnisse dieses Experiments.

6.4.1 Szene (a): Ländliches Gebiet bei Dresden

Nach Abbildung 6.1 sind hier die thematischen Klassen *Acker brach*, *Vegetation*, *Straße/Weg* und *Feld bewirtschaftet* gesucht. Diese Szene eignet sich aufgrund ihrer homogenen Flächen und einer einfachen und übersichtlichen Struktur sehr gut für eine anschauliche Darstellung und Interpretation der Eigenschaften der jeweils für die Klassifikation angewendeten Methodik. Aus diesem Grund wurde diese Szene auch vermehrt als Beispiel bei verschiedenen Prozessierungsschritten in dieser Arbeit angeführt. In Abbildung 6.4 ist die mittlere Gesamtklassifikationsgüte

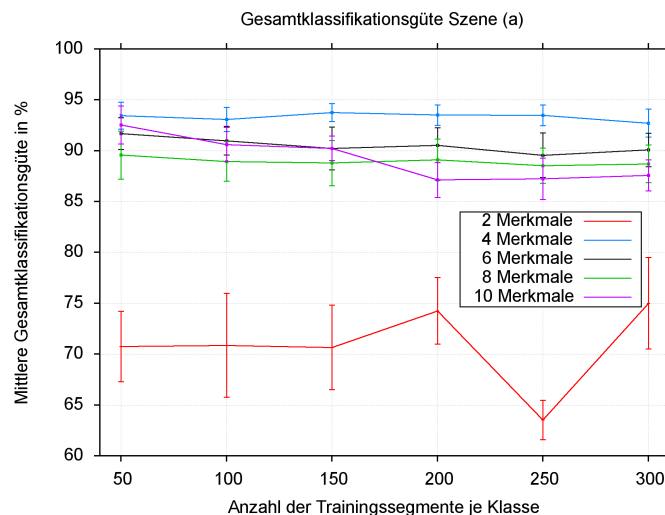


Abbildung 6.4: Gesamtklassifikationsgüten für Szene (a): Ländliches Gebiet bei Dresden. Die Fehlerbalken stellen den Wert 0.5σ dar.

te in Abhängigkeit der Anzahl der Trainingssegmente je Klasse für die unterschiedlichen Men-

gen von Merkmalen dargestellt. Diese Darstellung zeigt, dass hier aufgrund einer offensichtlich guten spektralen Trennbarkeit der Klassen schon mit einer einfachen ML-Klassifikation sehr gute Ergebnisse erzielt werden können. Während zwei Merkmale offensichtlich nicht für eine gute Trennung der Klassen ausreichen, sind für vier bis zehn Merkmale relativ konstant gute Ergebnisse mit den geschätzten Datenmodellen zu verzeichnen. Mit vier Merkmalen werden für alle Mengen von Trainingsdaten die besten Ergebnisse erzielt. Jede Hinzunahme weiterer Merkmale führt zu einer leichten Verschlechterung der mittleren Gesamtklassifikationsgüten, was vermutlich durch ein erhöhtes Rauschen durch die zusätzlichen redundanten Merkmale verursacht wird. Generell scheinen die Ergebnisse weitaus unabhängig von der Anzahl und Wahl der Trainingsdaten zu sein, was wiederum für eine hohe spektrale Homogenität der Klassen spricht.

Die Betrachtung der Konsumenten- und Produzentengenauigkeiten sowie die visuelle Beurteilung von Bild und Klassifikationsergebnissen (Abbildung 6.6) weisen darauf hin, dass die Klassen *Vegetation* und *Feld bewirtschaftet* ähnliche spektrale Eigenschaften haben. Dies äußert sich hauptsächlich durch eine Fehlklassifikation von Segmenten der Klasse *Feld bewirtschaftet* als *Vegetation*, was vor allem zu niedrigen Konsumentengenauigkeiten der Klasse *Vegetation* führt (Abbildung 6.5 links). Die Produzentengenauigkeiten der Klasse *Feld bewirtschaftet* (Abbildung 6.5 rechts) sind, vermutlich durch den Flächenanteil der Klasse von über 50% am Gesamtbild, hingegen nur sehr gering beeinflusst und ähneln vielmehr sehr den Gesamtklassifikationsgüten. Fehlklassifikationen äußern sich ansonsten nicht signifikant auffällig

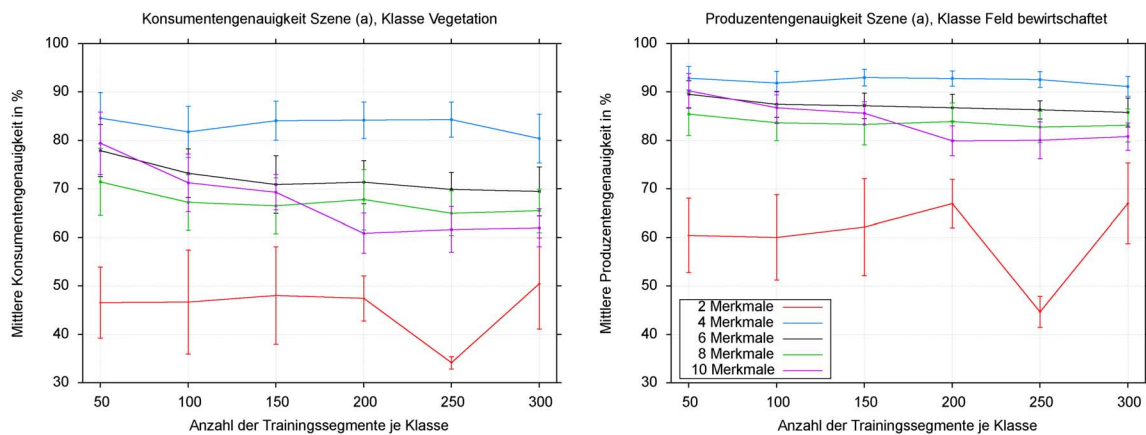


Abbildung 6.5: Konsumentengenauigkeiten der Klasse *Vegetation* (links) und Produzentengenauigkeiten der Klasse *Feld bewirtschaftet* (rechts) für Szene (a). Die Fehlerbalken stellen den Wert 0.5σ dar.

in den Klassifikationsgüten, sind jedoch aufgrund des übersichtlichen Bildinhalts sehr gut im visualisierten Klassifikationsergebnis zu erkennen. In Abbildung 6.6 ist ein beispielhaftes Ergebnis einer ML-Klassifikation mit einer Gesamtklassifikationsgüte von 93.0 % dargestellt. Einerseits fällt die oben beschriebene Konfusion zwischen den Klassen *Feld bewirtschaftet* und *Vegetation* auf. Zum anderen werden hier Fehlklassifikationen im Bereich von Klassenübergängen (Mischklassen) sowie vereinzelte Konfusionen zwischen den Klassen *Straße* und *Feld bewirtschaftet* deutlich.



Abbildung 6.6: Links: Referenz. Rechts: Ergebnis einer ML-Klassifikation mit Fehlklassifikationen im rechten Bildbereich sowie vereinzelt in Bereichen von Klassenübergängen.

6.4.2 Szene (b): Stadtgebiet von Dresden

Bei dieser Szene sollen die 2×2 Pixel großen Bildsegmente der untersten Ebene des Quadtree-Modells in die Klassen *Haus* und *Rest* klassifiziert werden. In Abbildung 6.7 sind die Ge-

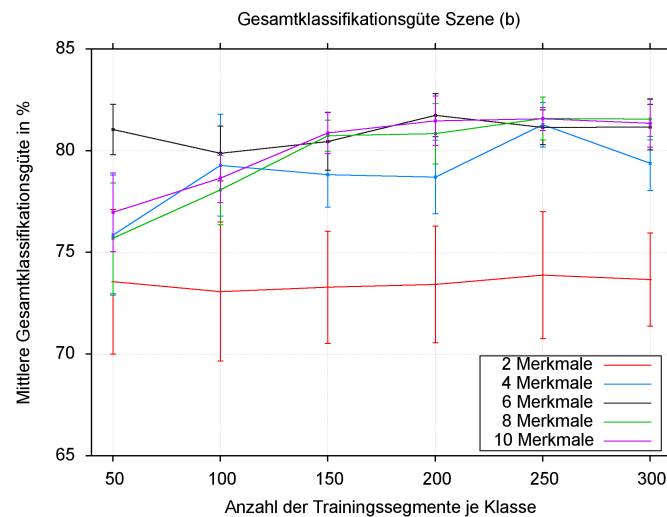


Abbildung 6.7: Gesamtclassifikationsgüten für Szene (b): Stadtgebiet Dresden. Die Fehlerbalken stellen den Wert 0.3σ dar.

samtclassifikationsgüten der ML-Klassifikationen für die unterschiedlichen Konfigurationen von Merkmalen und Trainingsdaten dargestellt. Unter der Verwendung von lediglich zwei Merkmalen werden auch hier konstant niedrige mittlere Klassifikationsgüten bei einer hohen Streuung σ erreicht. Bei einer Anzahl von vier bis zehn Merkmalen zeichnet sich mit einer wachsenden Zahl an Trainingssegmenten als allgemeiner Trend eine Konvergenz gegen eine

Gesamtklassifikationsgüte von ca. 81 % ab, wobei die Streuung der Ergebnisse abnimmt. Für ein optimales Schätzergebnis sollten hier demnach mindestens 200 Trainingssegmente je Klasse sowie mindestens vier Merkmale verwendet werden.

Die Konsumentengenauigkeit der Klasse *Haus* ist mit Werten im Bereich von 50 – 55 % sehr niedrig (Abbildung 6.8 rechts) und besagt, dass ungefähr die Hälfte aller als *Haus* klassifizierten Segmente tatsächlich aus der Menge der Klasse *Rest* stammen. Die Produzentengenauigkeiten in Abbildung 6.9 belegt dies durch die Aussage, dass im Mittel ca. 20 – 25 % der Testsegmente der Klasse *Rest* als *Haus* klassifiziert werden. Der Grund dafür liegt in der offensichtlich schlechten spektralen Trennbarkeit der Klasse *Haus* von den in der Klasse *Rest* beinhalteten Straßen, was durch einen visuellen Vergleich des Ergebnisses mit der Referenz (Abbildung 6.10) bestätigt wird. Die Klasse *Rest* hat hingegen durchgängig mittlere Konsumentengenauigkeiten

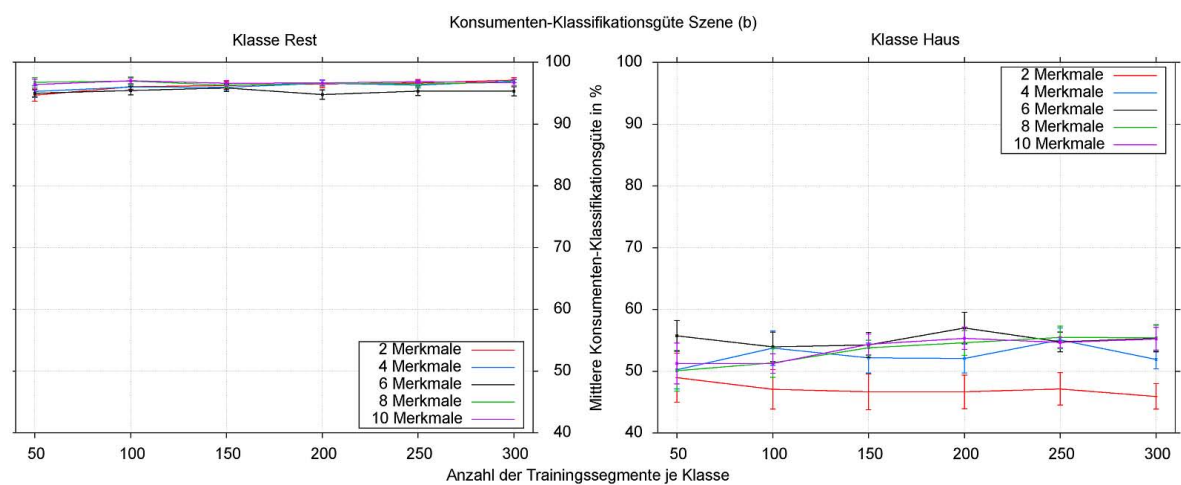


Abbildung 6.8: Konsumentengenauigkeiten der Klassen *Rest* und *Haus*. Die Fehlerbalken stellen den Wert 0.3σ dar.

von über 95 % (Abbildung 6.8 links). Insgesamt werden also viele Segmente der Klasse *Rest* als *Haus* klassifiziert, so dass die Konsumentengenauigkeit der Klasse *Rest* sehr hoch ist. Die Schwierigkeit bei diesem Klassifikationsproblem besteht demnach offensichtlich in der Trennung der Klasse *Haus* von den Straßen aufgrund von Hausdächern, welche vermutlich geteert sind.

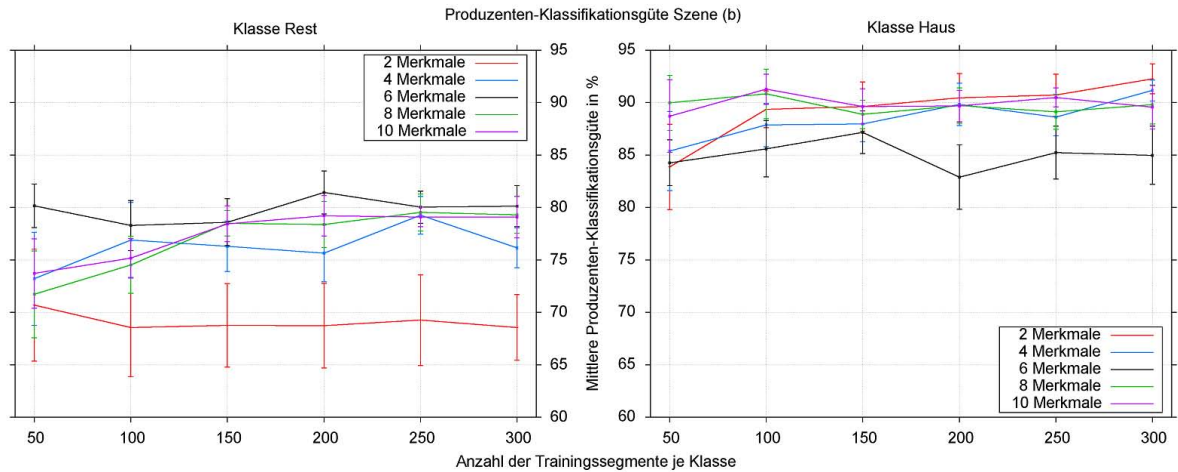


Abbildung 6.9: Produzentengenauigkeiten der Klassen *Haus* und *Rest*. Die Fehlerbalken stellen den Wert 0.3σ dar.

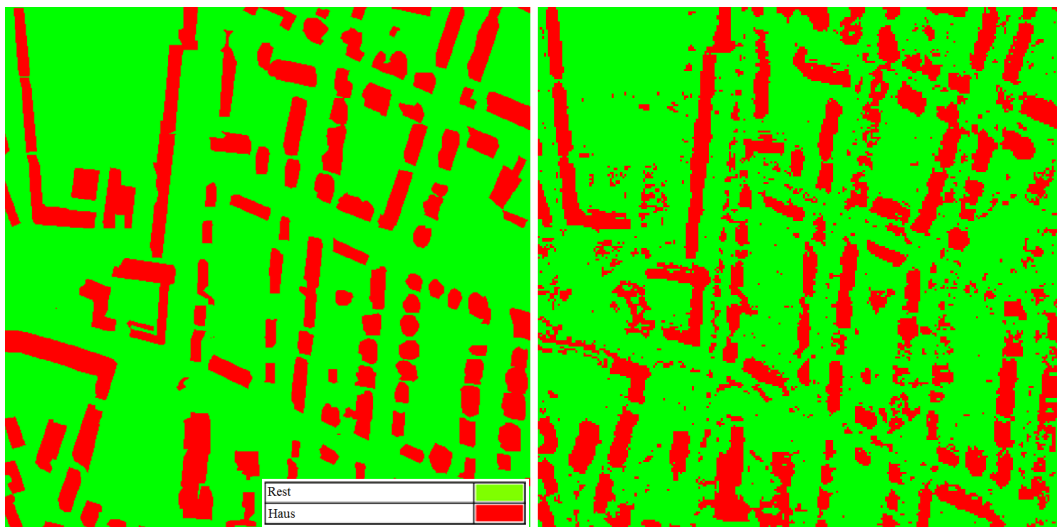


Abbildung 6.10: Ergebnis einer ML-Klassifikation mit einer Gesamtklassifikationsgüte von 87.1 %. Links: Referenz. Rechts: Ergebnis mit teilweise falsch klassifizierten Straßenbereichen und Hausdächern.

6.4.3 Szene (d): Ländlicher Raum mit Industrie

Anhand der Ergebnisse der Kreuzvalidierung bei Szene (d) mit den Zielklassen *Vegetation*, *Straße*, *Industrie* und *Freifläche* lässt sich ablesen, dass in diesem Experiment die Anzahl der Trainingsdaten und die Anzahl der Merkmale keinen signifikanten Einfluss auf die Qualität des Datenmodells haben. In Abbildung 6.11 lässt sich kein allgemeiner Trend ablesen. Der Grund für die recht großen Streuungen in den Klassifikationsgüten ist sehr wahrschein-

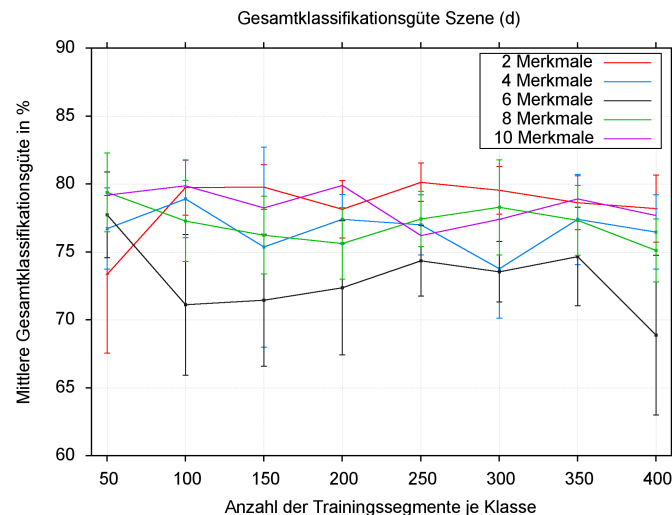


Abbildung 6.11: Ergebnisse für Szene (d): Ländlicher Raum mit Industrie. Die Fehlerbalken stellen den Wert 0.5σ dar.

lich durch die zufällige und daher offensichtlich teilweise nicht ausreichend repräsentative Wahl der Trainingsdaten bei spektral heterogenen und schwer voneinander trennbaren thematischen Klassen gegeben. Rein visuell wird diese Vermutung durch unterschiedliche Straßentypen, Industriebauten sowie mit dünner Vegetation bedecktem Freiboden bestätigt (Abbildung 6.2). Eine Kreuzvalidierung mit 20 zufällig gewählten Mengen von Trainingsdaten je Konfiguration könnte daher nicht ausreichend repräsentativ sein. Da eine Kreuzvalidierung sehr rechenintensiv und somit zeitaufwändig ist, erfolgt an dieser Stelle keine Berechnung mit weiteren Konfigurationen. Die Ergebnisse in den folgenden Abschnitten zeigen jedoch, dass eine sorgfältige Wahl der Trainingsdaten auch hier durchaus zu guten Klassifikationsergebnissen führt.

6.4.4 Szene (f): Stadtrand von Dujiangyan, Sichuan, China

Auch bei dieser Szene kann durch die Betrachtung der Gesamtklassifikationsgüten auf eine Unabhängigkeit dieser von der Anzahl der Trainingsdaten geschlossen werden. Auf die Klassifikation der Szene in die Klassen *Wasser*, *Siedlungsgebiet* und *Rest* hat lediglich die Anzahl der gewählten Merkmale einen geringen Einfluss, da mit vier Merkmalen die besten und mit weniger oder mehr Merkmalen leicht schlechtere Ergebnisse erreicht werden (vergleiche Abbildung 6.12). Konfusionen treten hier hauptsächlich zwischen den Klassen *Wasser* und *Siedlung* sowie zwischen *Siedlung* und *Rest* auf. Im ersten Fall werden Segmente der Klasse *Wasser* vermehrt als *Siedlung* klassifiziert, was zu niedrigen Produzentengenauigkeiten (Abbildung 6.13), jedoch zu hohen Konsumentengenauigkeiten für *Wasser* führt. Die Konsumentengenauigkeit

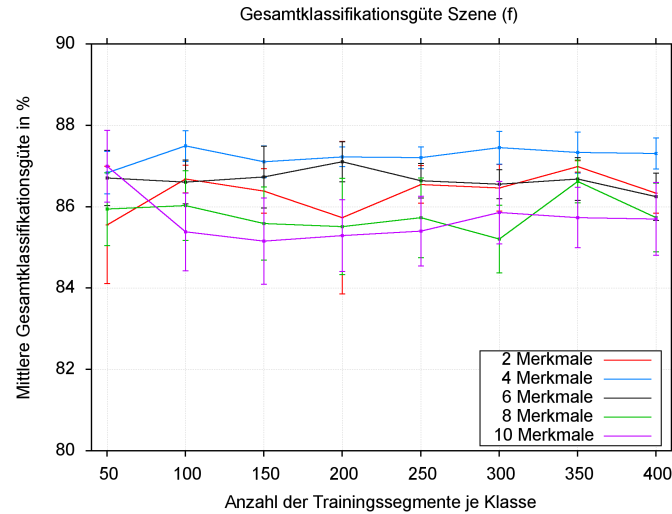


Abbildung 6.12: Gesamtklassifikationsgüten für Szene (f). Die Fehlerbalken stellen den Wert 0.5σ dar.

für die Klasse *Siedlung* wird dadurch und zusätzlich durch Fehlklassifikationen von *Rest* als *Siedlung* etwas verschlechtert, was jedoch durch den hohen Anteil an Siedlungsfläche an der Gesamtszene nicht signifikant in Abbildung 6.13 ersichtlich wird.

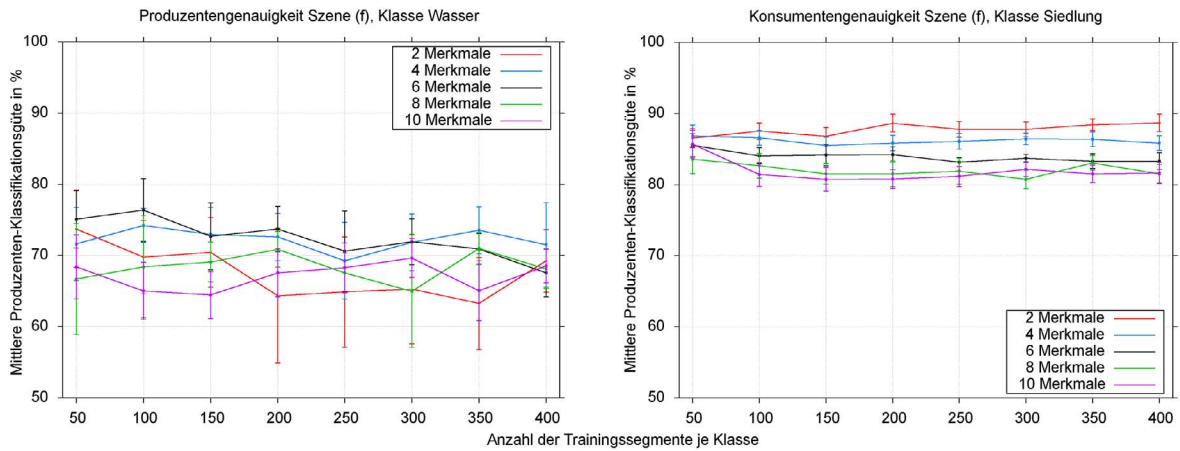


Abbildung 6.13: Links: Mittlere Produzentengenauigkeit der Klasse *Wasser*. Rechts: Mittlere Konsumentengenauigkeit der Klasse *Siedlung*. Die Fehlerbalken stellen den Wert 0.5σ dar.

6.4.5 Zusammenfassung

Die dargestellten Ergebnisse spiegeln die Schwierigkeit der semiautomatischen Klassifikation sowie die Komplexität von hochauflösenden Satellitenbildern wieder. Das wichtigste Ergebnis dieses Experiments ist, dass die Klassifikationsgüten weniger durch die Anzahl der Trainingssegmente oder der Merkmale, jedoch maßgeblich durch die Wahl der Trainingsdaten beeinflusst werden. Lediglich für Szene (b) zeigt sich ein Zusammenhang zwischen der Anzahl der Trai-

ningsdaten und den resultierenden Ergebnissen, wonach mindestens 200 Trainingssegmente je Klasse verwendet werden sollten. Durch die individuelle Selektion der relevanten Merkmale bei jeder Konfiguration sind jedoch in den meisten Fällen Gesamtklassifikationsgüten ohne eindeutigen Trend zu beobachten. Dieser Effekt wird bei spektral inhomogenen sowie schwer trennbaren Klassen vor allem bei den Szenen (c), (d) und (e) weiter verstärkt, was zu stark variierenden Klassifikationsgüten und hohen Varianzen dieser führt. Dies lässt darauf schließen, dass 20 zufällig gewählte Mengen an Trainingsdaten in diesen Fällen nicht ausreichen, um eine fundierte Aussage treffen zu können.

Generell kann konstatiert werden, dass für die obigen Testszenen in der Regel mit 200 bis 400 Trainingssegmenten je Klasse und einer Merkmalsanzahl größer gleich vier eine optimale Parameterschätzung möglich ist, wobei die Trainingsdaten sehr sorgfältig ausgewählt und ausreichend repräsentativ sein müssen. Es treten jedoch bei alleiniger Verwendung des Datenmodells mitunter starke Konfusionen zwischen den Klassen auf.

6.5 Eigenschaften der Parameterschätzung für das *a priori* Modell

In diesem Abschnitt wird die Robustheit der Schätzung der Parameter des *a priori* Modells in Abhängigkeit der verschiedenen Einflussfaktoren untersucht. Darüber hinaus soll mit diesem Versuchsaufbau eine erste Aussage über den Mehrwert des Einbezugs von hierarchischer Kontextinformation getroffen werden. Da die Parameterschätzung unabhängig vom Datenmodell erfolgt, wird auch hier jeweils ein festes Datenmodell angenommen, wobei die thematischen Klassen entsprechend Abschnitt 6.1 lediglich auf der feinsten Ebene des Quadrees modelliert werden. Etwaige Einflüsse von Datenmodellen anderer Skalen werden auf diese Weise vermieden. Die Einflussfaktoren bei der Schätzung der Übergangswahrscheinlichkeiten sind mit den Trainingsdaten sowie mit der Initialisierung der Übergangswahrscheinlichkeiten für den EM-Algorithmus gegeben. Daher erfolgt in diesem Experiment die Schätzung der Übergangswahrscheinlichkeiten bei unterschiedlichen Konfigurationen von Trainingsdaten und Initialisierungen. Für die Trainingsdaten werden auf der Grundlage der Referenz für jede thematische Klasse zufällig jeweils 20, 40, 60, 80 und 100 % aus der Menge aller Segmente gewählt. Die Übergangswahrscheinlichkeiten werden auf $k = 0, \dots, K$ unterschiedliche Arten initialisiert, wobei $K = 5$ gewählt wird. Es wird dabei jeweils ein Potts-Modell angenommen, dass auf den Hauptdiagonalen der Matrizen der Übergangswahrscheinlichkeiten durch

$$\theta = \frac{1}{C} + \left(1 - \frac{1}{C}\right) \cdot \frac{k}{K}, \quad (6.2)$$

mit der Anzahl der Klassen C , initialisiert wird. Die Werte abseits der Hauptdiagonalen werden jeweils durch $(1 - \theta)/(C - 1)$ initialisiert. Bei $k = 0$ ergibt sich somit eine Gleichverteilung, wohingegen bei $k = 5$ auf der Hauptdiagonalen $\theta = 1.0$ folgt. Hat eine Übergangswahrscheinlichkeit den Wert Null, so wird diese durch den EM-Algorithmus nicht mehr verändert [Feng et al., 2002]. Aus diesem Grund wird bei $k = 5$ auf der Hauptdiagonalen $\theta = 0.99$ definiert. Aus diesen Definitionen ergeben sich insgesamt 30 verschiedene Konfigurationen, mit denen für alle sechs Szenen eine Parameterschätzung durchgeführt wird. Als Bewertungskriterium für eine Konfiguration wird die Gesamtklassifikationsgüte der hierarchischen MPM-Inferenz herangezogen. In Abbildung 6.14 sind die Ergebnisse der Evaluierung für die sechs Szenen dargestellt. Für die sechs verschiedenen Initialisierungen sind die Gesamtklassifikationsgüten der hierarchischen Inferenz mit den geschätzten Parametern in Abhängigkeit der Menge der

Trainingssegmente in % von der Gesamtmenge dargestellt. Zum Vergleich ist zusätzlich das Ergebnis der ML-Klassifikation mit dem verwendeten Datenmodell angegeben.

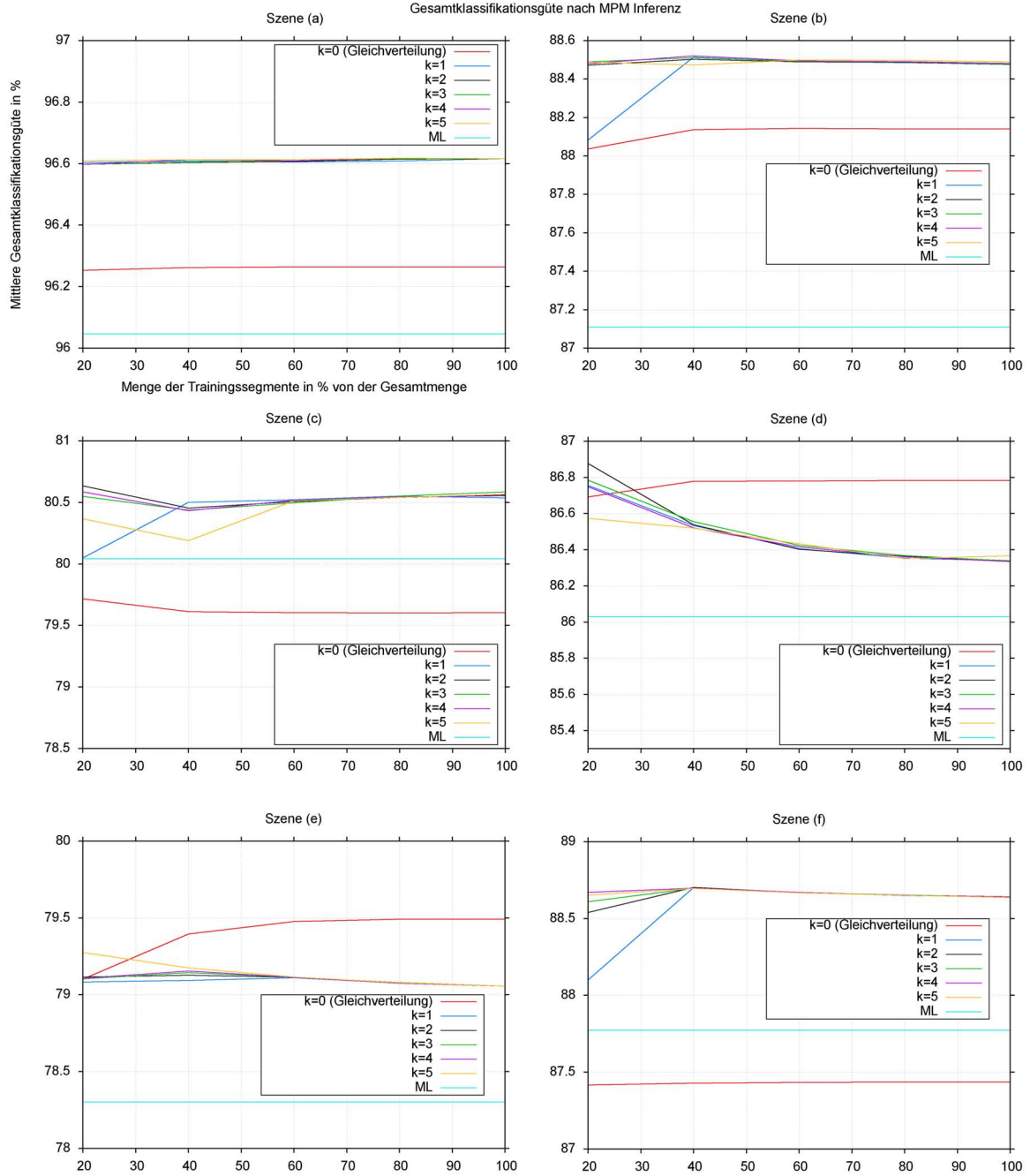


Abbildung 6.14: Ergebnisse der Evaluierung des *a priori* Modells für die Szenen (a) bis (f).

Die Ergebnisse des Experiments können wie folgt interpretiert werden:

Szene (a), ländliches Gebiet: Nach Abbildung 6.14 ist hier keine Abhängigkeit von der Anzahl der Trainingsdaten zu erkennen. Generell ergibt die EM-Schätzung für alle Konfigurationen $k \neq 0$ in den unteren Levels des Quadrees sehr hohe Werte auf den Hauptdiagonalen der Matrizen der Übergangswahrscheinlichkeiten, was auf die großen homogenen Flächen im Bild zurückzuführen ist. Die daraus resultierende Glättung durch die hierarchische Inferenz führt zu einer leichten Verbesserung der Klassifikationsgüte gegenüber der ML-Klassifikation. Bei einer Initialisierung der EM-Schätzung mit gleich verteilten Übergangswahrscheinlichkeiten ergeben sich keine sinnvollen Modellparameter. Diese liefern jedoch dennoch und vermutlich zufällig etwas bessere Ergebnisse als die ML-Klassifikation.

Szene (b), Stadtgebiet Dresden: Die Ergebnisse ähneln denen von Szene (a), jedoch ist bei einem Anteil von 20 % für die ersten beiden Konfigurationen eine leichte Verschlechterung im Inferenzergebnis zu beobachten. Dies kann auf die aus diesen Konfigurationen resultierenden vergleichsweise kleineren Übergangswahrscheinlichkeiten bei Klassengleichheit benachbarter Segmente zurückgeführt werden. Dieser Effekt entsteht durch die Annahme einer Gleichverteilung der Wahrscheinlichkeiten für die Klassenzugehörigkeit bei Segmenten, die nicht für das Training verwendet werden und verursacht eine Abschwächung der Glättung bei der Inferenz. Bei einer Initialisierung mit größeren Werten auf den Hauptdiagonalen sind bei wenigen Trainingsdaten entsprechend größere Werte im Schätzergebnis zu beobachten, während beispielsweise bei 100 % Trainingsdatenanteil für die Konfigurationen $k \neq 0$ fast identische Schätzergebnisse folgen. Daraus kann abgeleitet werden, dass die Initialisierung für wenige Trainingsdaten einen höheren Einfluss auf das Schätzergebnis hat. Die Inferenzergebnisse sind auch hier stets besser, als bei einer ML-Klassifikation.

Szene (c), Stadtgebiet Dresden mit Wasser: Die Initialisierung mit einer Gleichverteilung führt bei dieser Szene aufgrund eines nicht gewünschten lokalen Maximums bei der EM-Schätzung zu schlechteren Inferenzergebnissen als bei einer ML-Klassifikation. Bei einem Trainingsdatenanteil von 20 % sind für $k \neq 0$ teilweise die besten Inferenzergebnisse zu verzeichnen, was wiederum durch den oben beschriebenen Effekt der abgeschwächten Glättung beziehungsweise Favorisierung von Klassengleichheiten benachbarter Segmente begründet werden kann. Während bei $k = 2$ mit 100 % der Trainingsdaten im untersten Level Übergangswahrscheinlichkeiten auf der Hauptdiagonale von $\theta_{100} = [0.98997, 0.92674, 0.980880]$ geschätzt werden, lauten die entsprechenden Werte bei 20 % $\theta_{20} = [0.82595, 0.86458, 0.87919]$. Die entsprechenden Inferenzergebnisse unterscheiden sich quantitativ jedoch nur um 0.1 %.

Szenen (d) und (e), ländlicher Raum mit Industrie und Siedlungen: Hier sind die Ergebnisse mit der hierarchischen Inferenz ebenfalls besser als mit der ML-Klassifikation. Das ähnliche Verhalten bei den Ergebnissen der Inferenz in Abbildung 6.14 könnte aus der ähnlichen räumlichen Struktur der Bildinhalte beziehungsweise der thematischen Klassen resultieren. In beiden Szenen sind kleine Strukturen, wie beispielsweise Straßen, Gebäude und Siedlungsbereiche enthalten. Gerade in solchen Fällen kann eine übermäßige Glättung durch sehr hohe Übergangswahrscheinlichkeiten bei Klassengleichheit zu einer Vernachlässigung dieser feinen Strukturen führen. Die geschätzten Übergangswahrscheinlichkeiten führen hier offensichtlich bei großen Mengen von Trainingsdaten zu einer zu starken Glättung, so dass anzunehmen ist,

dass diese Strukturen respektive Klassenübergänge verhältnismäßig zu wenig in den Trainingsdaten vorkommen. Eine Verwendung von sehr wenigen Trainingsdaten (20 %) führt wiederum, zu einer verstärkten Abschwächung des Glättungseffekts. Auch hier sind die quantitativen Unterschiede sehr gering.

Szene (f), Stadtrand in China: Die Ergebnisse dieser Szene sind vergleichbar mit denen von Szene (b). Die Initialisierung der EM-Schätzung mit einer Gleichverteilung führt hier sogar zu schlechteren Inferenzergebnissen, als bei einer ML-Klassifikation.

Auf der Grundlage dieses Experiments kann geschlussfolgert werden, dass durchaus eine Abhängigkeit zwischen der Menge der Trainingsdaten für die EM-Schätzung und der Initialisierung der Übergangswahrscheinlichkeiten einerseits und den daraus resultierenden Modellparametern des *a priori* Modells andererseits besteht. Während die Bereitstellung der Trainingsdaten für die Schätzung der Parameter des Datenmodells manuell erfolgt, wird an dieser Stelle die Trainingsdatenwahl durch die Wahl des Fehlerniveaus α im Rahmen der ML-Vorklassifikation reguliert. Erfolgt die Wahl von α derart, dass mindestens 40 % aller Segmente als Trainingsdaten verwendet werden, so ergeben sich für unterschiedliche Initialisierungen kaum Unterschiede in den korrespondierenden Inferenzergebnissen. Die Übergangswahrscheinlichkeiten bezüglich gleicher Klassen in benachbarten Hierarchieebenen sind in diesen Fällen sehr hoch, wodurch das Modell zu einer Glättung der Inferenzergebnisse neigt. Die Szenen (d) und (e) weisen feine und linienhafte Strukturen auf, die in diesen Fällen im Rahmen der Inferenz mitunter vernachlässigt werden. Bei einer Verwendung von sehr wenigen Trainingsdaten (durch ein entsprechend hohes Fehlerniveau) wird diesem Effekt entgegengewirkt, da für alle Segmente, die keine Klassenzugehörigkeit haben (also keine Trainingssegmente sind), gleiche Wahrscheinlichkeiten für jede Klasse angenommen werden. Dieser Effekt wird wiederum besonders in Szenen mit kleinen Strukturen deutlich und führt zu einem Erhalt dieser. Daher sollte sich die Auswahl von Trainingsdaten und (demnach die Wahl des Fehlerniveaus α bei der ML-Vorklassifikation) für die EM-Schätzung an der räumlichen Struktur der thematischen Klassen orientieren. Bei feinen Strukturen, wie beispielsweise Straßen, welche vom Bildmodell lediglich durch Segmente auf der feinsten Ebene des Quadrees repräsentiert werden können, sollten demnach weniger Trainingsdaten (ein hohes Fehlerniveau) verwendet werden, während bei großen homogenen Flächen eine verstärkte Robustheit gegenüber Initialisierung und Trainingsdatenmenge zu beobachten ist. Die Wahl der Trainingsdaten, wie sie in diesem Experiment erfolgt, ist aufgrund der fehlenden Referenz üblicherweise nicht möglich. Im folgenden Abschnitt wird daher der Zusammenhang zwischen der Bereitstellung von Trainingsdaten für die EM-Schätzung auf der Grundlage des Datenmodells und den daraus resultierenden Inferenzergebnissen untersucht.

Die Ergebnisse der hierarchischen MPM-Inferenz sind quantitativ immer besser als eine lokale ML-Klassifikation, sofern die Initialisierung der Übergangswahrscheinlichkeiten für die EM-Schätzung ungleich einer Gleichverteilung ist. Der quantitative Mehrwert der hierarchischen Inferenz gegenüber einer ML-Klassifikation zeigt sich mit maximalen Verbesserungen der Gesamtklassifikationsgüten von ca. 1 % in diesem Experiment nicht signifikant.

6.6 Bereitstellung von Trainingsdaten für das *a priori* Modell

Die Bereitstellung von Trainingsdaten in Form von klassifizierten Bildsegmenten erfolgt auf der Grundlage einer beschränkten ML-Klassifikation (vergleiche Abschnitt 4.3). Die Entschei-

dung, ob ein Bildsegment mit dem jeweiligen Datenmodell eindeutig und mit einem definierten Fehlerniveau α klassifiziert werden kann, wird anhand eines $\chi^2_{\alpha,f}$ -Tests getroffen. In diesem Abschnitt soll gezeigt werden, inwieweit mit dieser Herangehensweise tatsächlich Fehlklassifikationen vermieden werden können und ob eine allgemeine Aussage beziehungsweise Empfehlung für die Definition des Fehlerniveaus α getroffen werden kann. Die bisher gewonnenen Erkenntnisse bezüglich des *a priori* Modells zeigen, dass vor allem bei räumlich klein strukturierten thematischen Klassen weniger Trainingsdaten zu besseren Ergebnissen führen, was durch die daraus resultierende Abschwächung des Glättungseffekts bei der hierarchischen Inferenz begründet werden kann. Für alle Szenen werden im Folgenden die Gesamtfehlerraten GF der ML-Klassifikation mit $\chi^2_{\alpha,f}$ -Test sowie für die entsprechenden Gesamtklassifikationsgüten GK der hierarchischen MPM-Inferenz in Abhängigkeit vom Fehlerniveau α tabellarisch dargestellt. Zu Vergleichszwecken sind in den letzten beiden Zeilen die Ergebnisse bei einer unbeschränkten ML-Klassifikation sowie die besten Ergebnisse der MPM-Inferenz aus Abschnitt 6.5 angegeben. Für die EM-Schätzung der Übergangswahrscheinlichkeiten werden hier alle klassifizierten Segmente verwendet. Eine steigendes Fehlerniveau α entspricht immer kleine-

α	Szene											
	(a)		(b)		(c)		(d)		(e)		(f)	
	GF	GK	GF	GK	GF	GK	GF	GK	GF	GK	GF	GK
0.1	1.25	96.55	5.34	88.43	9.72	80.43	4.18	86.52	5.95	78.80	5.99	88.66
0.2	0.88	96.59	6.32	88.43	11.31	80.46	5.64	86.45	6.42	78.81	7.36	88.64
0.3	0.71	96.60	7.24	88.43	12.15	80.49	7.07	86.42	7.03	78.81	8.27	88.62
0.4	0.61	96.63	7.56	88.43	12.87	80.52	7.91	86.45	8.22	78.82	8.70	88.62
0.5	0.52	96.64	7.59	88.41	13.06	80.53	8.07	86.51	10.29	78.83	8.71	88.62
0.6	0.43	96.67	6.89	88.42	12.97	80.54	7.86	86.51	13.23	78.85	8.29	88.63
0.7	0.38	96.67	5.74	88.42	12.91	80.60	7.18	86.52	17.00	78.87	7.53	88.66
0.8	0.31	96.73	3.84	88.48	13.05	80.65	6.16	86.50	22.55	78.92	6.47	88.67
0.9	0.23	96.66	1.71	88.60	12.88	80.71	4.36	86.58	20.35	78.86	5.58	88.70
<i>ML</i>	3.95	96.53	12.89	88.41	19.96	80.71	13.97	86.30	21.70	78.72	12.23	88.53
<i>max_{MPM}</i>	-	96.62	-	88.53	-	80.64	-	86.88	-	79.49	-	88.70

Tabelle 6.3: Gesamtfehlerraten GF der beschränkten ML-Klassifikation sowie Gesamtklassifikationsgüten GK der hierarchischen MPM-Inferenz nach EM-Schätzung der *a priori* Modellparameter in Abhängigkeit vom Fehlerniveau α .

ren Ellipsoiden gleicher Wahrscheinlichkeiten $P = 1 - \alpha$ um die Komponentenzentren. Für Szene (a) ist dieser Zusammenhang zwischen einem wachsendem Wert für α und einem sich entsprechend verringernden Gesamtfehler GF bei der ML-Klassifikation gut ablesbar. In Abbildung 6.15 sind einige Ergebnisse dieser ML-Klassifikation dargestellt. Für die schwarzen Bildsegmente wurde die Nullhypothese verworfen. Bei allen anderen Szenen ist mit $\alpha = 0.1$ der Gesamtfehler bei der ML-Klassifikation zunächst signifikant kleiner als ohne Test, steigt jedoch mit größer werdenden Fehlerniveaus zunächst an und fällt mit $\alpha \rightarrow 1.0$ wieder ab. Dieses Verhalten ist auf die Nullhypothese (4.9), bei der für einen Punkt im Merkmalsraum (respektive einem Merkmalsvektor) neben dem Vergleich der Mahalanobis-Distanzen mit den Quantilen $\chi^2_{\alpha,f}$ der χ^2 -Verteilung auch die Eindeutigkeit der Klassenzugehörigkeit geprüft wird, zurückzuführen. Bei kleinen Fehlerniveaus sind entsprechend große Ellipsoide und somit auch verstärkt Mehrdeutigkeiten in der Zuordnung eines Punktes zu den Klassen zu erwarten. Wird α größer, so sind immer mehr eindeutige Zuordnungen zu erwarten, wobei diese offensichtlich dennoch falsch sein können.

Die minimalen Fehlerraten der beschränkten ML-Klassifikation sind bei $\alpha = 0.1$ und $\alpha = 0.2$ sowie $\alpha = 0.9$ zu beobachten (hellgraue Felder in Tabelle 6.3). Die besten Inferenzergebnisse wiederum (dunkelgraue Felder in Tabelle 6.3) werden mit einem *a priori* Modell, geschätzt

aus dem Ergebnis einer ML-Klassifikation mit $\alpha = 0.8$ und $\alpha = 0.9$, erzielt. Die absoluten Unterschiede in den Inferenzergebnissen sind jedoch sehr gering und unterscheiden sich um maximal 0.28 % (Szene (d)). Des Weiteren werden hier ähnliche Ergebnisse, wie mit der MPM-Schätzung (\max_{MPM}) erreicht. Anhand der Ergebnisse lässt sich ableiten, dass das Fehlniveau α für die Bereitstellung von Trainingsdaten möglichst hoch gewählt werden sollte. Es ist zu beachten, dass in diesem Fall mitunter nur sehr wenige Trainingsdaten vorhanden sind, wodurch nach eigenen Erfahrungen ein Maximum bei der EM-Schätzung in der Regel nur mit einer vergleichsweise langsamen Konvergenz gefunden wird.

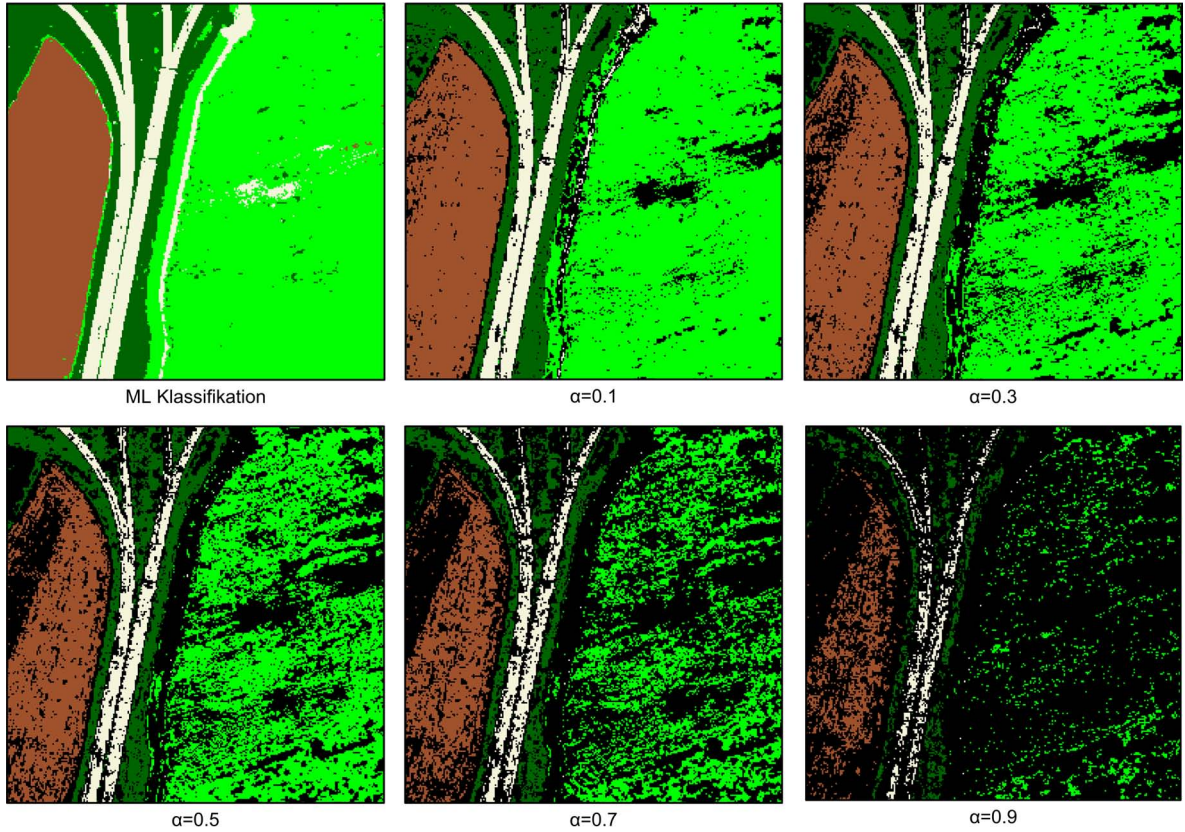


Abbildung 6.15: Trainingsdaten für die Schätzung der Parameter der *a priori* Modells in Abhängigkeit des Fehlniveaus α

6.7 Multiskalenprozessierung

In diesem Abschnitt wird die Methode hinsichtlich des Mehrwerts der Prozessierung unter Verwendung von Bildinformation auf mehreren Skalen S untersucht. Dazu werden für jede Szene Datenmodelle in den untersten drei Hierarchieebenen des Quadrees (S^N, S^{N-1}, S^{N-2}) trainiert und bei der hierarchischen Inferenz miteinander kombiniert. Die Auswahl genau dieser Konfigurationen begründet sich durch die Erkenntnisse in [Yang et al., 2009], wonach mit mehr als drei Hierarchieebenen keine Verbesserungen in den Ergebnissen zu verzeichnen sind. Insgesamt werden die folgenden vier Konfigurationen betrachtet: (S^N), (S^N, S^{N-1}), (S^N, S^{N-1}, S^{N-2}) und (S^N, S^{N-2}), wobei mit einer minimalen Segmentgröße von 2×2 Pixel

für die kleinen Szenen $N = 8$ und für die größeren Szenen $N = 9$ gilt. Sofern es möglich ist, werden jeweils in allen Ebenen die in Abschnitt 6.1 beschriebenen thematischen Klassen modelliert. Bei Szene (f) ist bei einer Segmentgröße von 16×16 Pixel (Level S^7) eine Erfassung der Wasserflächen aufgrund ihrer schmalen Form nicht mehr möglich, so dass in diesem Level die Klassen *Siedlungsflächen* und *Rest* modelliert werden. Des Weiteren werden in Szene (e) in Level S^7 aus gleichem Grund lediglich die Klassen *Wald* und *Rest* modelliert.

Basierend auf den Ergebnissen in Abschnitt 6.5 wird die EM-Schätzung der Parameter des *a priori* Modells mit $\theta = 0.75$ für Klassengleichheit initialisiert. Für das Training der Datenmodelle werden jeweils etwa 200 Trainingssegmente je Klasse sowie vier Merkmale verwendet (Minimalanforderungen laut Abschnitt 6.4). In den Tabellen 6.4 und 6.6 sind die Gesamtklassifikationsgüten der MPM-Inferenz für die verschiedenen Konfigurationen zusammen gestellt, wobei das beste Ergebnis jeweils grau gekennzeichnet ist. Anhand der Gesamtklassifikations-

Szene	Gesamtklassifikationsgüte in %				
	ML	(S^8)	(S^8, S^7)	(S^8, S^7, S^6)	(S^8, S^6)
(a)	96.05	96.73	96.76	96.75	96.68
(b)	87.11	88.60	88.79	88.76	88.65
(c)	80.04	80.71	81.89	81.93	81.12

Tabelle 6.4: Gesamtklassifikationsgüten der hierarchischen MPM-Inferenz der Szenen (a), (b) und (c) bei verschiedenen Kombinationen von Skalen. Die grau hinterlegten Felder kennzeichnen das jeweils beste Ergebnis.

güten lässt sich für die Szenen (a), (b) und (c) zunächst kein großer Mehrwert der Multiskalenprozessierung ablesen. Anhand eines detaillierteren Vergleichs der Konsumenten- und Produzentengenauigkeiten (UA und PA) aus lokaler ML-Klassifikation und der besten MPM-Inferenz in Tabelle 6.5 lassen sich die Unterschiede besser verdeutlichen. Der Großteil dieser

Szene	Klasse	ML-Klassifikation		Beste MPM-Inferenz	
		UA in %	PA in %	UA in %	PA in %
(a)	<i>Acker brach</i>	99.73	97.80	99.88	97.71
	<i>Vegetation</i>	95.62	89.08	98.74	89.02
	<i>Strasse/Weg</i>	81.48	96.95	83.09	97.05
	<i>Feld bewirtschaftet</i>	97.69	97.38	97.03	98.72
(b)	<i>Haus</i>	67.19	79.59	73.71	75.32
	<i>Rest</i>	94.02	89.20	93.10	92.54
(c)	<i>Wasser</i>	99.32	73.24	99.68	79.88
	<i>Haus/vers. Fläche</i>	71.31	77.15	74.19	78.93
	<i>Rest</i>	81.94	84.59	82.97	84.88

Tabelle 6.5: Gegenüberstellung der prozentualen Konsumenten- (UA) und Produzentengenauigkeiten (PA) aus ML-Klassifikation und der besten hierarchischen MPM-Inferenz für die Szenen (a), (b) und (c). Die grau hinterlegten Felder kennzeichnen eine Verbesserung durch die MPM-Inferenz.

Genauigkeitsmaße hat sich mit der hierarchischen Inferenz teilweise signifikant verbessert, was eine Verringerung der Konfusionen zwischen den thematischen Klassen belegt. In Szene (b) ist

bei der Klasse *Haus* mit 6.52 % die größte Verbesserung der Konsumentengenauigkeit, bei einer gleichzeitig verbesserten Produzentengenauigkeit für die Klasse *Rest*, zu verzeichnen, woraus folgt, dass weniger Segmente der Klasse *Rest* falsch als Haus klassifiziert werden. Umgekehrt werden jedoch vermehrt (dunkel erscheinende) Häuser als *Rest* (Straße) klassifiziert, so dass man hier auch von einer Verlagerung der Konfusionen sprechen kann. Die größte Verbesserung

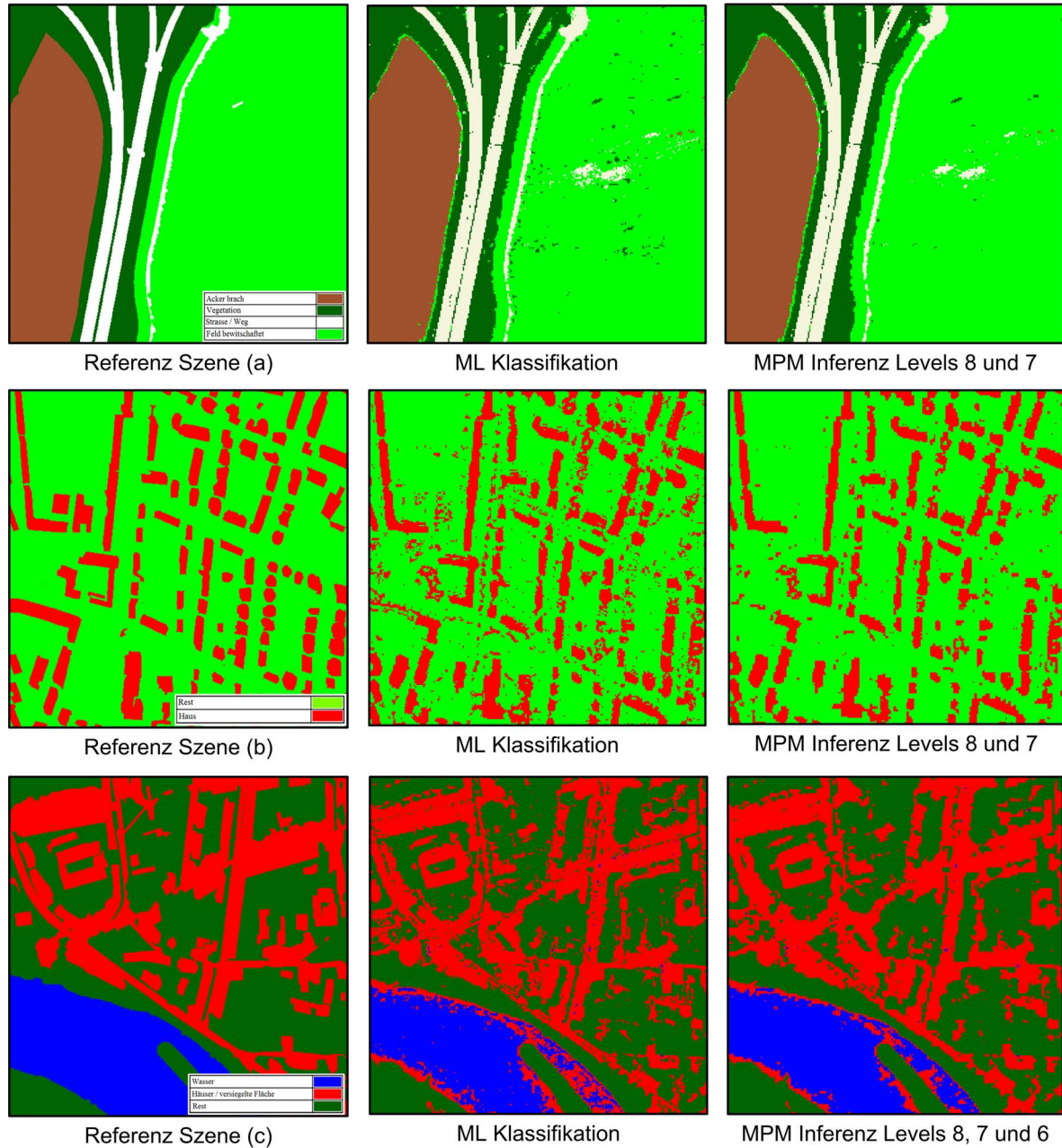


Abbildung 6.16: Visuelle Gegenüberstellung der Ergebnisse aus ML-Klassifikation und MPM-Inferenz für die Szenen (a), (b) und (c).

bezüglich der Produzentengenauigkeit ist mit 6.64 % bei der Klasse *Wasser* in Szene (c) zu verzeichnen. Im Zuge der ML-Klassifikation werden Segmente der Klasse *Wasser* vermehrt falsch als *Künstliche Struktur* klassifiziert. Mit der MPM-Inferenz werden mehr Referenzsegmente der

Klasse *Wasser* auch als solches klassifiziert, wodurch die Konsumentengenauigkeit der Klasse *Häuser/versiegelte Fläche* um 2.88 % verbessert wird. In Szene (a) fallen die Verbesserungen quantitativ geringer aus, jedoch werden die Unterschiede der Ergebnisse rein visuell in allen drei Szenen deutlich (Abbildung 6.16). Generell ist ein Glättungseffekt in den Ergebnissen zu beobachten, was dazu führt, dass in Szene (a) das rechte Feld homogener und in Szene (b) die vereinzelt Fehlklassifikationen von Segmenten als *Haus* verringert werden. In Szene (c) bestätigt die Abbildung die Verbesserung der Produzentengenauigkeit der Klasse *Wasser*. Die hierarchische Inferenz bewirkt zwar eine leichte Verringerung der Fehlklassifikationen, jedoch können diese beispielsweise in Bereichen von Klassenübergängen nicht vermieden werden.

Die erreichten Gesamtklassifikationsgüten bei der hierarchischen Inferenz für die Szenen (d), (e) und (f) sind in Tabelle 6.6 zusammengestellt. Auch hier fällt bei der Betrachtung der

Szene	Gesamtklassifikationsgüte in %				
	<i>ML</i>	(S^9)	(S^9, S^8)	(S^9, S^8, S^7)	(S^9, S^7)
(d)	86.03	86.58	87.01	87.17	86.84
(e)	78.30	78.92	79.18	79.25	79.88
(f)	87.38	88.70	89.14	89.0	87.68

Tabelle 6.6: Gesamtklassifikationsgüten der hierarchischen MPM-Inferenz der Szenen (d), (e) und (f) bei verschiedenen Kombinationen von Skalen. Die grau hinterlegten Felder kennzeichnen das jeweils beste Ergebnis.

Gesamtklassifikationsgüten auf, dass sich zwar ein quantitativer Mehrwert durch die Multiskalenprozessierung ergibt, dieser jedoch mit maximal 1.76 % gegenüber einer ML-Klassifikation mit dem gleichen Datenmodell nur sehr gering ist. Eine Gegenüberstellung der Konsumenten- und Produzentengenauigkeiten in Tabelle 6.7 verdeutlicht jedoch, dass auch hier die Mehrdeutigkeiten zwischen den thematischen Klassen im Merkmalsraum partiell verringert werden können. In Szene (d) werden beispielsweise mit der Kombination (S^9, S^8, S^7) die Produzenten-

Szene	Klasse	ML-Klassifikation		Beste MPM-Inferenz	
		UA in %	PA in %	UA in %	PA in %
(d)	<i>Vegetation</i>	96.24	91.09	96.26	90.45
	<i>Strasse</i>	61.18	75.66	63.02	77.60
	<i>Industrie</i>	70.66	83.74	76.51	85.36
	<i>Freifläche</i>	81.45	66.58	83.43	79.71
(e)	<i>Wald</i>	95.58	87.76	95.53	88.19
	<i>Künstliche Strukturen</i>	39.35	72.07	41.91	70.21
	<i>Rest</i>	82.96	71.03	83.55	74.60
(f)	<i>Wasser</i>	93.01	79.79	96.26	78.68
	<i>Siedlung</i>	85.52	94.21	86.74	95.24
	<i>Rest</i>	91.53	79.54	92.83	81.46

Tabelle 6.7: Gegenüberstellung der prozentualen Konsumenten- (UA) und Produzentengenauigkeiten (PA) aus ML-Klassifikation und bester hierarchischer MPM-Inferenz für die Szenen (d), (e) und (f). Die grau hinterlegten Felder kennzeichnen eine Verbesserung durch die MPM-Inferenz.

und Konsumentengenauigkeit der Klasse *Freifläche* um 13.13 % respektive 1.98 % verbessert, was auch in Abbildung 6.17 gut zu erkennen ist. Für fast alle thematischen Klassen bringt die Multiskalenprozessierung Verbesserungen der Klassifikationsgüten mit sich.

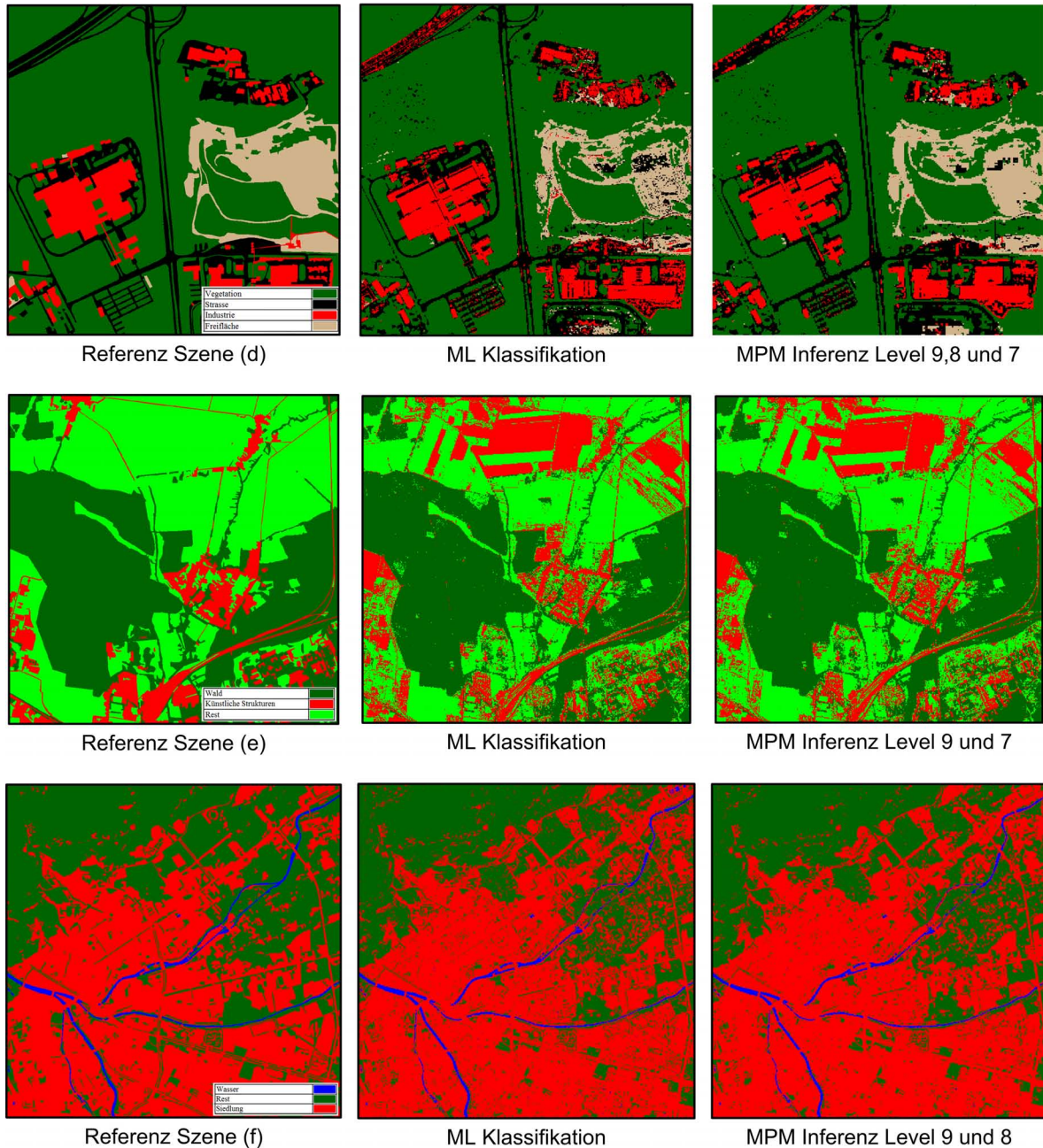


Abbildung 6.17: Visuelle Gegenüberstellung der Ergebnisse aus ML-Klassifikation und MPM-Inferenz für die Szenen (d), (e) und (f).

Die Gegenüberstellung der verschiedenen Ergebnisse zeigt, dass mit dem Einbezug von hierarchischer Kontextinformation sowie von Bildinformation auf mehreren Skalen prinzipiell Konfusionen zwischen thematischen Klassen verringert werden können. Dies kann an den leicht

verbesserten Gesamtklassifikationsgüten gegenüber einer lokalen ML-Klassifikation abgelesen werden, verdeutlicht sich jedoch merklich bei der Betrachtung der Konsumenten- und Produzentengenauigkeiten sowie bei einem visuellen Vergleich der Ergebnisse. Hauptsächlich werden durch das hierarchische Modell Fehlklassifikationen in homogenen Flächen deutlich verringert (Glättung), was beispielsweise in den Ergebnissen von Szene (a) sehr gut zu erkennen ist. In Szene (b) werden Mehrdeutigkeiten verringert und deutlich weniger Straßenbereiche falsch als *Haus* klassifiziert. Jedoch ist festzuhalten, dass dadurch wiederum auch einige Häuser falsch als *Rest* klassifiziert werden. An allen Ergebnissen und vor allem an Szene (e) wird jedoch deutlich, dass sehr viele Fehlklassifikationen auch durch den Einbezug von mehreren Hierarchieebenen nicht vermieden werden können. Zum einen führt die starre Quadtree-Struktur mit seinen quadratischen Bildelementen dazu, dass Segmente mit Mischklassen sehr wahrscheinlich in jedem Bild auftreten. Auch unter Einbezug von hierarchischer Kontextinformation ist es daher offensichtlich wahrscheinlich, dass falsche Entscheidungen aufgrund des Datenmodells für solche Mischsegmente zu Fehlklassifikationen führen. Diese Überlegung wird durch Erfahrungen mit der Methode sowie durch die hier dargestellten Ergebnisse bestätigt. Ein Beispiel dafür ist mit dem Randbereich des linken Feldes in Szene (a) gegeben (Abbildung 6.16). Zum anderen können mitunter großflächige Fehlklassifikationen mit dem ursprünglichen hierarchischen Modell nicht vermieden werden. In Szene (e) werden große Flächen falsch als *Künstliche Strukturen* klassifiziert, was sich einerseits in sehr geringen Konsumentengenauigkeiten dieser Klasse ablesen lässt und andererseits durch Abbildung 6.17 bestätigt wird.

Insgesamt kann anhand der Erkenntnisse in diesem Abschnitt festgestellt werden, dass mit dem sehr starren Quadtree-Bildmodell für diese sechs Szenen im Zuge der hierarchischen Inferenz überwiegend durch Glättung leichte Verbesserungen in den Klassifikationsgüten gegenüber einer ML-Klassifikation erreicht werden. Großflächige Fehlklassifikationen, wie sie auch unter alleiniger Verwendung des Datenmodells auftreten, können jedoch offensichtlich mit diesem Modell nicht vermieden werden. In [Kersten et al., 2010] wurde anhand einer QuickBird-Szene beispielhaft gezeigt, dass die Multiskalenprozessierung mit diesem Modell durchaus großflächige Fehlklassifikationen verringern kann. Des Weiteren konnten mit einer modifizierten hierarchischen MAP-Inferenz Fehlklassifikationen von Mischsegmenten, beispielsweise in Bereichen von Klassenübergängen, stark verringert werden. Im folgenden Abschnitt wird untersucht, inwieweit die Auswirkungen aufgrund der starren Quadtree-Struktur durch das hybride Modell, respektive den Einbezug von räumlichem Kontext, einerseits sowie durch die modifizierten Inferenzmethoden mICM und mMPM andererseits abgeschwächt werden können. Diese Ergebnisse werden zusätzlich mit denen aus zwei weiteren Klassifikationsmethoden verglichen.

6.8 Vergleich der Methode mit anderen gängigen Verfahren

Nach der Evaluierung der Trainingsprozeduren für Daten- und *a priori* Modell sowie bezüglich des Mehrwerts der hierarchischen Inferenz nach [Laferté et al., 2000] folgt nun der praktische Vergleich des hybriden Modells mit anderen gängigen Klassifikationsmethoden. Insbesondere soll an dieser Stelle auch der Mehrwert des hybriden Modells, respektive der Kombination von hierarchischem und räumlichem Kontext, sowie der modifizierten Inferenzmethoden mMPM und mICM gegenüber den ursprünglichen Methoden herausgestellt werden.

Generell bietet sich ein Vergleich mit Methoden an, die thematisch verwandt und als *state of the art* anerkannt sind. Eine Gegenüberstellung von Verfahren der hierarchischen graphenbasierten Bildmodellierung ist in der Vergangenheit nur sehr vereinzelt, beispielsweise in [Awasthi et al., 2007], erfolgt, was unter anderem mit den sehr verschiedenen Anwendungsgebieten begründet werden kann. Eine weitere Barriere für einen fairen Vergleich ist die sehr begrenzte Verbreitung und Verfügbarkeit von Methoden aus diesem Themengebiet. Insbesondere erscheint neben einem theoretischen Vergleich von generativen und diskriminativen Methoden auch ein praktischer Vergleich dieser sehr interessant, ist jedoch im Rahmen dieser Arbeit aufgrund der oben dargelegten Gründe nicht zustande gekommen. An einem Vergleich der Methode mit den in Abschnitt 3.2 behandelten diskriminativen TCRFs wird jedoch momentan gearbeitet. Im Rahmen von weiterführenden Untersuchungen ist die Fertigstellung und Publikation dieses Vergleichs geplant.

In dieser Arbeit erfolgt daher ein Vergleich des hybriden Markov-Modells mit einer überwachten und pixelbasierten ML-Klassifikation sowie mit dem in [Bouman und Shapiro, 1994] vorgestellten Multiskalen Zufallsfeld (MSRF). Die überwachte ML-Klassifikation erfolgt in ERDAS IMAGINE Version 9.3, wobei alle vier jeweils vorliegenden spektralen Kanäle eines Bildes verwendet werden. Darüber hinaus werden auf das Ergebnis dieser pixelbasierten Klassifikation die Operatoren *clump* und *eliminate* bezüglich einer Nachbarschaft zweiter Ordnung angewendet.

Das Multiskalen Zufallsfeld (MSRF) basiert, wie auch der in dieser Arbeit verfolgte Ansatz, auf einer Quadtree-Struktur, bei der die Knoten (Segmente) in den einzelnen Hierarchieebenen als eine *top-down* Markov-Kette interpretiert werden. Die feinste Ebene des Quadtrees ist hier mit dem Bild selbst gegeben. Es handelt sich um einen generativen Ansatz, bei dem die thematischen Klassen in der feinsten Ebene mittels GMMs modelliert werden. Bei der MAP-Inferenz wird üblicherweise eine Kostenfunktion angesetzt, die Klassifikationsfehler unabhängig von der Skala des Quadtrees bestraft. Daraus resultieren bei einer Fehlklassifikation der Wurzel des Baumes (diese repräsentiert das gesamte Bild) die gleichen Kosten, wie bei der Fehlklassifikation eines einzelnen Pixels. Die in [Bouman und Shapiro, 1994] vorgeschlagene Kostenfunktion berücksichtigt die jeweilige Skala, so dass größere Kosten bei Fehlklassifikationen von großen Segmenten entstehen. Die daraus resultierende Inferenzmethode wird sequentielle MAP-Inferenz (SMAP) genannt. Durch die Anwendung eines Potts-Modells wird die Gleichheit benachbarter Pixel favorisiert. Die Ausprägung der Glättung durch die SMAP-Inferenz ist abhängig vom Verhalten der Bilddaten. Wenn die Daten vermitteln, dass benachbarte Pixel oft die Klasse wechseln, so wird der Glättungsgrad verringert [GRASS Team, 2011]. Insgesamt wird in feineren Auflösungen eine größere Wahrscheinlichkeit von Klassengleichheit benachbarter Bildelemente vorausgesetzt. Die benötigten Parameter für die Formulierung der skalenabhängigen Übergangswahrscheinlichkeiten werden während der Inferenz geschätzt, so dass hier insgesamt zwei komplette Durchläufe durch den Quadtree erforderlich sind. Der Ansatz ist in einem frei verfügbaren geographischen Informationssystem (GIS) mit dem Namen

geographical resources analysis support system (GRASS) ab der Version 4.1 integriert. Der Vergleich erfolgt mit der Version 6.4.

In Tabelle 6.8 sind zunächst die erreichten Gesamtklassifikationsgüten der drei verschiedenen Methoden für alle Szenen dargestellt. Alle Klassifikationen basieren auf sorgfältig und manuell ausgewählten Trainingsdaten. Für die hierarchische Inferenz mit dem hybriden Modell werden die Trainingsdaten für das Datenmodell aus Abschnitt 6.7 verwendet. Die Parameterschätzung für das *a priori* Modell basiert auf der EM-Schätzung unter Verwendung der Ergebnisse einer beschränkten ML-Klassifikationen. Dabei wird jeweils das in Abschnitt 6.6 identifizierte Fehlerniveau α gewählt. Das hybride Modell bietet eine Vielzahl verschiedener Prozessierungsmöglichkeiten. Insbesondere werden bei diesem Vergleich die folgenden Berechnungen durchgeführt und gegenüber gestellt:

- ML: *Maximum-likelihood*-Klassifikation in ERDAS mit anschließender Nachbearbeitung durch die Operatoren *clump* und *eliminate* bezüglich einer Nachbarschaft zweiter Ordnung.
- SMAP: Sequentielle MAP-Inferenz mit GRASS nach [Bouman und Shapiro, 1994].
- MPM: Hierarchische MPM-Inferenz mit dem hybriden Modell. Es wird hier jeweils das beste Ergebnis aus der Untersuchung bezüglich der Multiskalenprozessierung herangezogen. Der Parameter α ist das Fehlerniveau für den $\chi^2_{f,\alpha}$ -Test im Zuge der Trainingsdatenbereitstellung für das *a priori* Modell.
- MPM-ICM: Hierarchische MPM-Inferenz und anschließender Einbezug von räumlichem Kontext durch den ICM-Algorithmus mit dem hybriden Modell. Der Parameter β ist dabei das Gewicht des Kontextterms.
- MPM-mICM: Anwendung des modifizierten ICM-Algorithmus (Abschnitt 4.7) auf das Ergebnis der MPM-Inferenz im hybriden Modell. Der Grenzwert für die Entropie wird hier in allen Fällen $T_H = 0.0001$ gesetzt.
- mMPM: Modifizierte MPM-Inferenz nach Abschnitt 4.5.3.
- mMPM-mICM: Anwendung des modifizierten ICM-Algorithmus auf des Ergebnis der modifizierten MPM-Inferenz.

Die Klassifikationsgüten für die ML- und die SMAP-Klassifikation wurden auf der Grundlage von 3000 zufällig und entsprechend der prozentualen Flächenanteile der Klassen in der Referenz verteilten Punkten ermittelt. Die Gesamtklassifikationsgüten in Tabelle 6.8 sowie die Konsumenten- und Produzentengenauigkeiten in Tabelle 6.9 offenbaren, dass mit dem hybriden Modell bei fünf der sechs Szenen signifikant bessere Ergebnisse als mit der ML-Klassifikation und der SMAP-Inferenz erreicht werden. Es fällt dabei auf, dass die jeweils besten Gesamtklassifikationsgüten durch verschiedene Prozessierungswege erreicht werden, wobei die quantitativen Unterschiede größtenteils sehr gering ausfallen. Eine Anwendung der Methoden mMPM und mICM führt bei drei der sechs Szenen zum besten Ergebnis. In den anderen drei Fällen ist die maximale Differenz dieser Kombination zu dem besten Ergebnis des hybriden Modells mit 0.17 % sehr gering. Unter Verwendung des hybriden Modells sowie der modifizierten Inferenzmethoden fällt ein teilweise signifikanter Genauigkeitsgewinn gegenüber der ursprünglichen hierarchischen MPM-Inferenz auf. Bei Szene (e) ist dieser mit 7.5 %

Szene	Parameter	Gesamtklassifikationsgüten in %						
		ML ERDAS	SMAP GRASS	MPM	MPM ICM	MPM mICM	mMPM	mMPM mICM
(a)	$\alpha = 0.8$ $\beta = 1000$	97.13	96.70	96.76	97.31	97.15	96.96	97.54
(b)	$\alpha = 0.9$ $\beta = 50$	81.90	83.60	88.79	89.64	89.70	88.62	89.60
(c)	$\alpha = 0.9$ $\beta = 20$	82.00	77.47	81.93	83.19	82.94	81.44	83.02
(d)	$\alpha = 0.9$ $\beta = 5.0$	89.87	83.07	87.32	87.75	87.77	87.67	88.18
(e)	$\alpha = 0.8$ $\beta = 5.0$	82.40	82.40	79.88	81.09	80.61	82.36	87.38
(f)	$\alpha = 0.9$ $\beta = 2.0$	83.87	79.47	89.14	89.45	89.48	89.63	90.03

Tabelle 6.8: Gegenüberstellung der Gesamtklassifikationsgüten aus ML-Klassifikation, SMAP-Inferenz und der vorgestellten Methode. Die grau hinterlegten Felder kennzeichnen das beste Ergebnis.

am höchsten. Es ist an dieser Stelle zu bemerken, dass das Ergebnis der modifizierten MPM-Inferenz lediglich als Zwischenergebnis respektive Eingangskonfiguration bei der ICM-Inferenz dienen sollte, da hier das Auftreten von Blockstrukturen im Klassifikationsergebnis mitunter stark begünstigt wird. Es folgt eine detaillierte Darstellung der Ergebnisse für die Szenen (a), (b), (d), (e) und (f).

Szene	Klasse	ML ERDAS		SMAP GRASS		Bestes Ergebnis hybrides Modell	
		UA in %	PA in %	UA in %	PA in %	UA in %	PA in %
(a)	<i>Acker brach</i> <i>Vegetation</i> <i>Strasse/Weg</i> <i>Feld bew.</i>	99.83	93.93	100.00	93.61	99.88	99.25
		98.52	92.46	94.20	93.45	98.62	90.56
		89.31	97.74	91.23	98.11	85.59	93.82
		97.17	99.75	97.26	98.70	98.52	99.65
(b)	<i>Haus</i> <i>Rest</i>	58.40	82.47	62.68	74.43	77.64	73.91
		93.87	81.73	91.68	86.41	92.84	94.08
(c)	<i>Wasser</i> <i>Künstl. Strukt.</i> <i>Rest</i>	99.12	83.09	98.92	73.02	99.97	81.23
		72.04	87.78	62.64	90.84	76.21	80.38
		87.62	78.20	89.66	69.89	83.80	85.96
(d)	<i>Vegetation</i> <i>Strasse</i> <i>Industrie</i> <i>Freifläche</i>	91.63	97.21	83.36	96.98	95.86	91.01
		83.65	73.73	93.75	39.82	65.68	79.20
		93.44	75.50	86.49	86.49	80.88	86.04
		81.13	77.48	64.71	49.25	85.12	83.72
(e)	<i>Wald</i> <i>Künstl. Strukt.</i> <i>Rest</i>	97.87	81.45	96.80	86.62	95.98	89.48
		68.28	60.19	53.57	50.00	64.73	70.90
		76.19	91.02	78.39	86.99	86.44	89.79
(f)	<i>Wasser</i> <i>Siedlung</i> <i>Rest</i>	100.00	39.39	100.00	28.57	96.41.	78.98
		79.96	92.37	73.31	98.04	87.04.	96.70.
		88.33	87.98	95.98	58.41	94.91.	81.63

Tabelle 6.9: Gegenüberstellung der prozentualen Konsumenten- (UA) und Produzentengenauigkeiten (PA) aus ML-Klassifikation in ERDAS, SMAP-Inferenz in GRASS und bestem Ergebnis mit dem hybriden Modell für alle Szenen. Die grau hinterlegten Felder kennzeichnen jeweils Verbesserungen mit dem hybriden Modell (UA: hellgrau, PA: dunkelgrau).

6.8.1 Ergebnisse für Szene (a)

In Szene (a) fallen bei ML und SMAP verhältnismäßig niedrige Konsumentengenauigkeiten der Klasse *Strasse* (89.31 % und 91.23 %) sowie niedrige Produzentengenauigkeiten der Klasse *Vegetation* (92.46 % und 93.45 %) auf (Tabelle 6.9), was durch eine partielle Aussparung des Weges sowie einer leichten Verbreiterung der Strasse in Bereiche der Vegetation zurückzuführen ist. Bei dem Ergebnis der SMAP-Inferenz sind zudem verstärkt Fehlklassifikationen des bewirtschafteten Feldes als *Vegetation* zu verzeichnen (vergleiche Abbildung 6.18 und Konsumentengenauigkeit *Vegetation*: 94.20 %). Des Weiteren treten sowohl bei ML als auch bei SMAP Fehlklassifikationen im Randbereich der Klasse *Acker brach* auf, was bei dieser Klasse

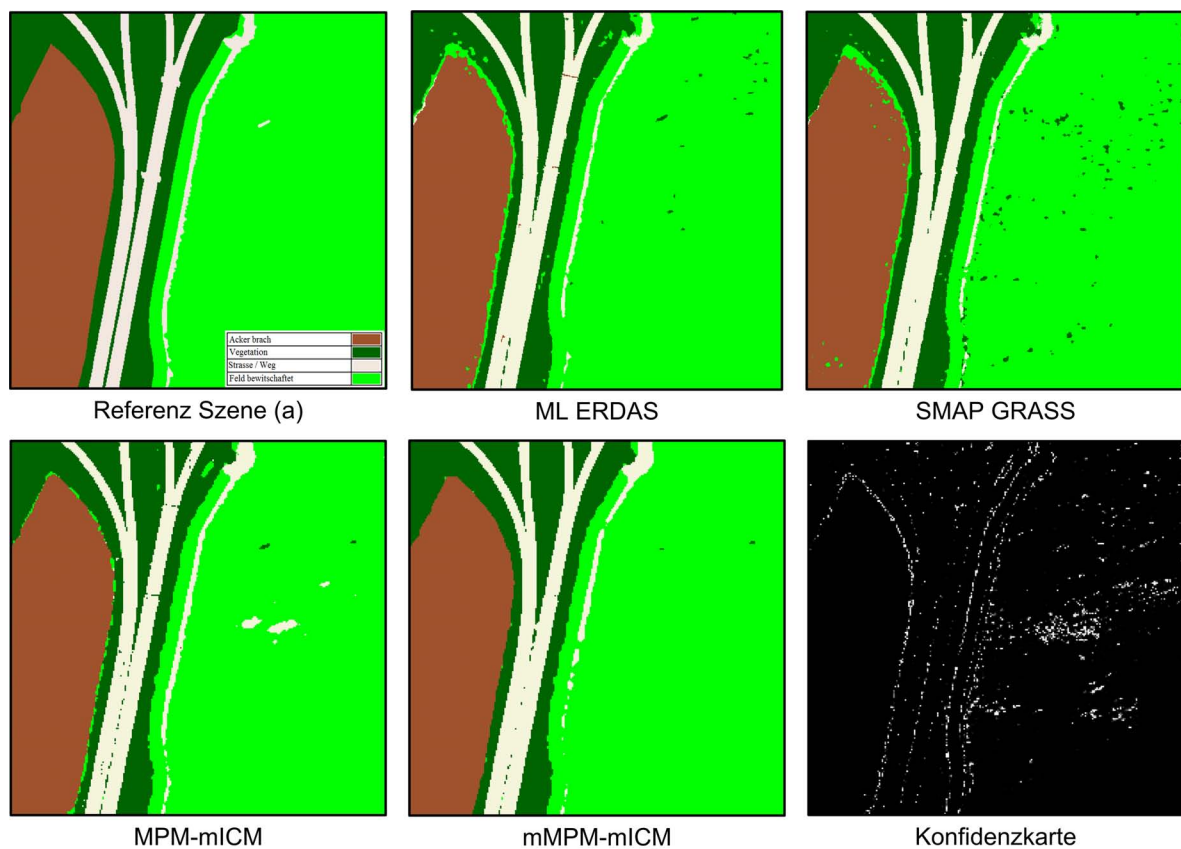


Abbildung 6.18: Konfidenzkarte und Ergebnisse der Methoden ML, SMAP, MPM-mICM, mMPM-mICM für Szene (a).

zu niedrigeren Produzentengenauigkeiten gegenüber dem besten Ergebnis aus dem hybriden Modell führt (93.93 % (ML) und 93.61 % (SMAP) versus 99.25 % (mMPM-mICM)). Die Konfidenzkarte weist auf nicht eindeutig klassifizierbare Bildsegmente in den Randbereichen der thematischen Klassen hin. Die Fehlklassifikationen innerhalb homogener Flächen sowie in Randbereichen von thematischen Flächen können durch die modifizierte hierarchische Inferenz stark verringert werden. Das Ergebnis aus der Kombination der modifizierten Inferenzmethoden eliminiert Fehlklassifikationen im Randbereich der Klasse *Feld bewirtschaftet* fast komplett, was anhand der leicht verbesserten Konsumentengenauigkeit in Tabelle 6.9 zu sehen ist.

Auch wenn das Ergebnis des hybriden Modells qualitativ sehr gut aussieht und laut Tabelle

6.8 die beste Gesamtklassifikationsgüte aufweist, fallen hier eine verschlechterte Konsumentengenauigkeit der Klasse *Strasse/Weg* sowie eine verschlechterte Produzentengenauigkeit der Klasse *Vegetation* auf. Dies ist auf eine verstärkte Vernachlässigung beziehungsweise Glättung des Weges sowie auf die Verbreiterung der Strasse in Bereiche der Vegetation zurückzuführen.

6.8.2 Ergebnisse für Szene (b)

Bei Szene (b) sind mit ML und SMAP die Konfusionen zwischen den beiden Klassen *Haus* und *Rest* anhand der vergleichsweise geringen Produzentengenauigkeiten beider Klassen zu beobachten. Die damit einhergehende sehr niedrige Konsumentengenauigkeit der Klasse *Haus* zeigt, dass sich diese Klasse vermehrt aus Segmenten der Klasse *Rest* zusammen setzt, was auf die spektrale Ähnlichkeit zwischen dunklen Hausdächern und Straßen zurückzuführen ist. Dies

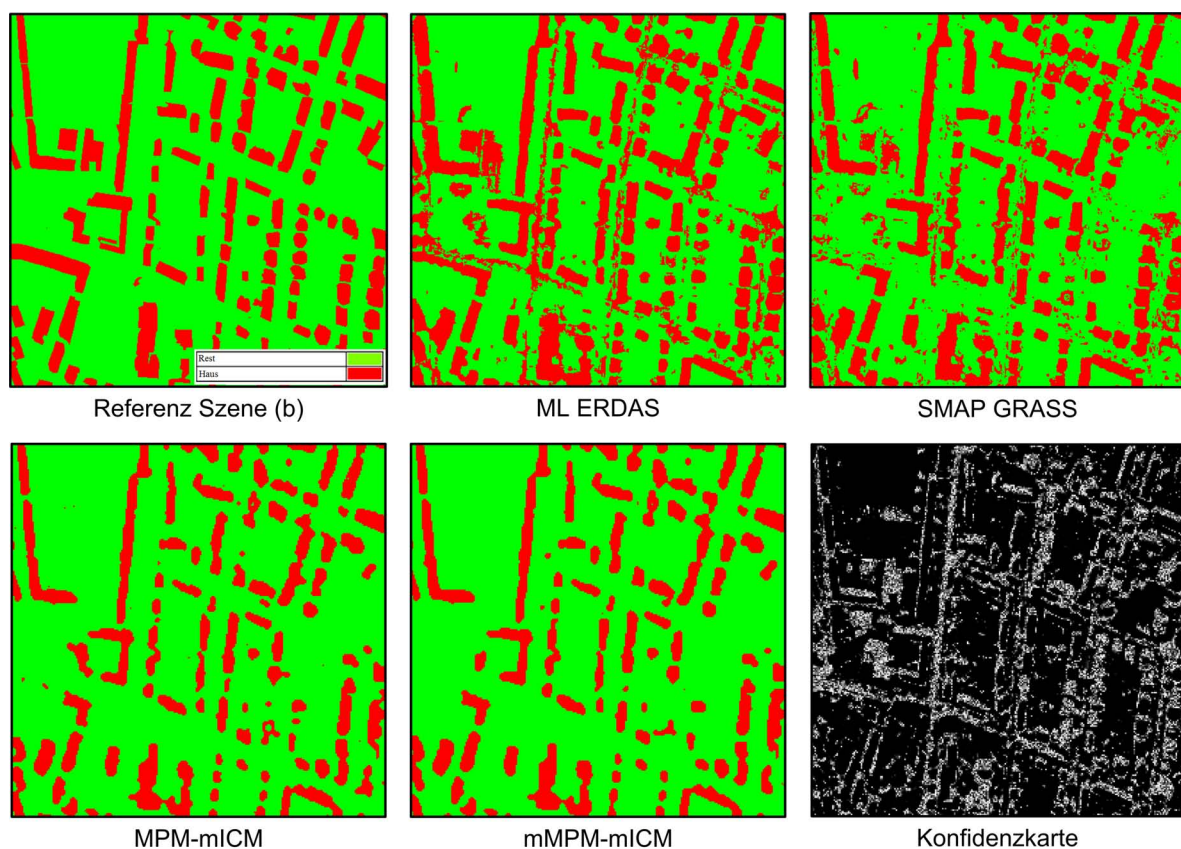


Abbildung 6.19: Konfidenzkarte und Ergebnisse der Methoden ML, SMAP, MPM-mICM, mMPM-mICM für Szene (b).

bestätigt sich in Abbildung 6.19, wo für die Verfahren ML und SMAP in erster Linie offensichtliche Fehlklassifikationen von Straßenbereichen als *Haus* zu erkennen sind. Die Schwierigkeit der spektralen Trennbarkeit bei dieser Szene zeigt sich zudem auch in der Konfidenzkarte der MPM-Inferenz deutlich. Durch das hybride Modell (MPM-mICM und mMPM-mICM) können die Fehlklassifikationen von Straßenbereichen als Haus stark verringert werden, was sich signifikant in höheren Konsumentengenauigkeiten der Klasse *Haus* sowie Produzentengenauigkeiten der Klasse *Rest* wieder spiegelt. Der Anteil der Fehlklassifikationen von Segmenten der

Klasse *Haus* wird hingegen sogar Verstärkt (Produzentengenauigkeit von 73.91 % mit MPM-ICM versus 82.74 % mit ML). Insgesamt werden also mit dem hybriden Modell die Straßen nicht mehr falsch als *Haus*, jedoch umgekehrt einige Häuser ebenfalls als *Rest* klassifiziert. Ein Vergleich der Klassifikationsergebnisse mit dem Bild (Abbildung 6.1) bestätigt, dass es sich dabei um dunkle Hausdächer handelt. Die Konfusionen werden demnach zwar mit den modifizierten Inferenzmethoden verringert jedoch kann man auch von einer Verlagerung dieser sprechen. Die Ergebnisse MPM-mICM und mMPM-mICM unterscheiden sich visuell und quantitativ kaum.

6.8.3 Ergebnisse für Szene (d)

Bei Szene (d) stellt das Ergebnis der ML-Klassifikation mit einer Gesamtklassifikationsgüte von 89.87 % das beste Ergebnis dar, wohingegen mit der hybriden mMPM-mICM Inferenz 88.18 % erreicht werden. In Abbildung 6.20 sowie in Tabelle 6.9 sind jedoch deutliche Unter-

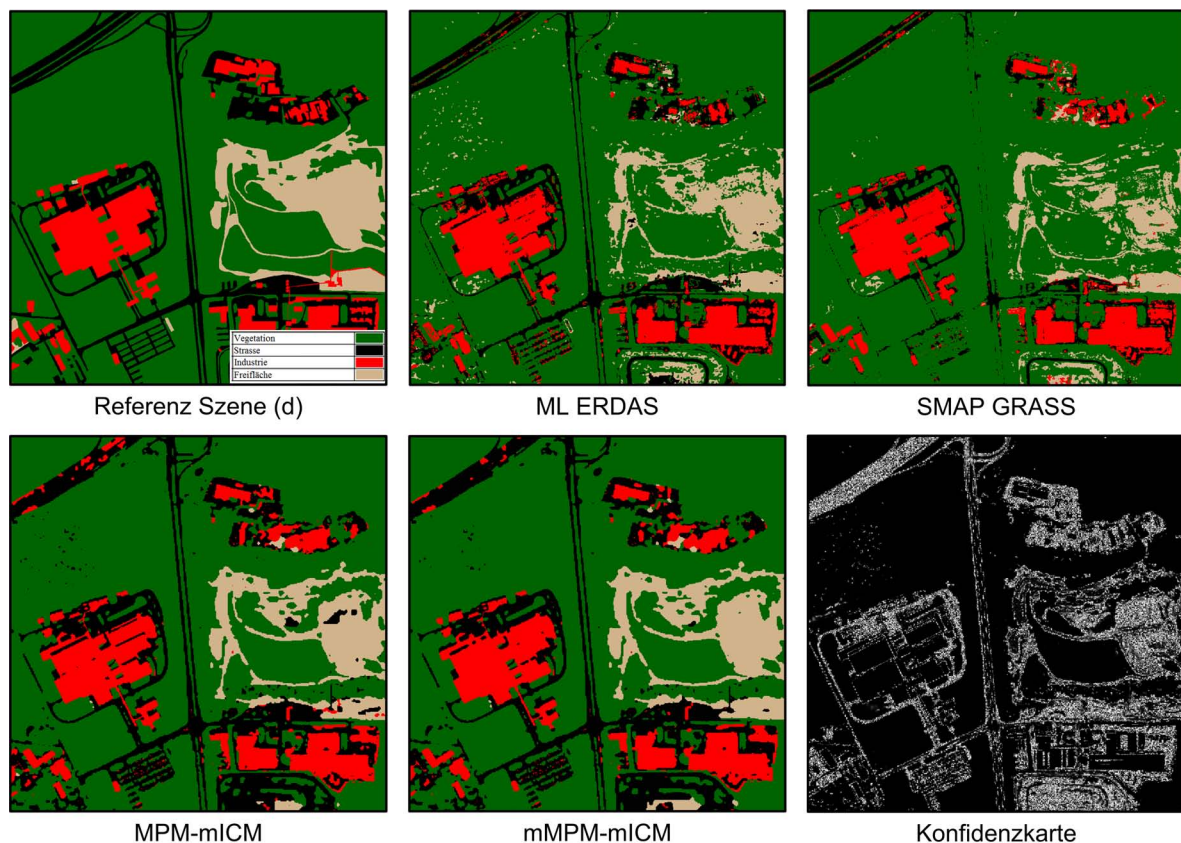


Abbildung 6.20: Konfidenzkarte und Ergebnisse der Methoden ML, SMAP, MPM-mICM, mMPM-mICM für Szene (d).

schiede zwischen den Klassifikationen zu erkennen. Die SMAP-Inferenz führt zu einer starken Glättung, so dass die Freiflächenbereiche sowie feine Strukturen, wie Straßen, teilweise verschwinden (Produzentengenauigkeiten von 39.82 % und 49.25 %). Mit dem hybriden Modell werden die Freiflächen sowie die Straßen vergleichsweise besser erhalten, was sich in den entsprechenden maximalen Produzentengenauigkeiten von 79.20 % und 83.72 % ablesen lässt.

Konfusionen zwischen den Klassen *Industrie* und *Strasse* können jedoch auch hier nicht vermieden werden. Die Konsumentengenauigkeit der Klasse *Strasse* verschlechtert sich dadurch sowie durch Konfusionen mit der Klasse *Vegetation* sogar signifikant gegenüber ML und SMAP. Die Konfidenzkarte verdeutlicht auch hier die Schwierigkeit der Trennbarkeit zwischen allen Klassen, ausgenommen von *Vegetation*.

6.8.4 Ergebnisse für Szene (e)

Anhand der Ergebnisse für Szene (e) zeigt sich der Mehrwert des hybriden Modells sehr deutlich. Eine Anwendung der Verfahren ML und SMAP führt zu Konfusionen zwischen den Klassen *Künstliche Strukturen* und *Rest*, was zu vergleichsweise geringen Konsumentengenauigkeiten (ML: 68.28 %, SMAP: 53.57 %) sowie Produzentengenauigkeiten (ML: 60.19 %, SMAP: 50.00 %) der Klasse *Künstliche Strukturen* führt. Die Multiskalenprozessierung (hierarchische MPM-Inferenz) sowie die Kombination von MPM- und ICM-Inferenz verstärkt die großräumigen Fehlklassifikationen sogar weiter, was in Tabelle 6.8 an den Gesamtklassifikationsgüten sowie in Abbildung 6.21 zu erkennen ist. Im Zuge der modifizierten MPM-Inferenz un-

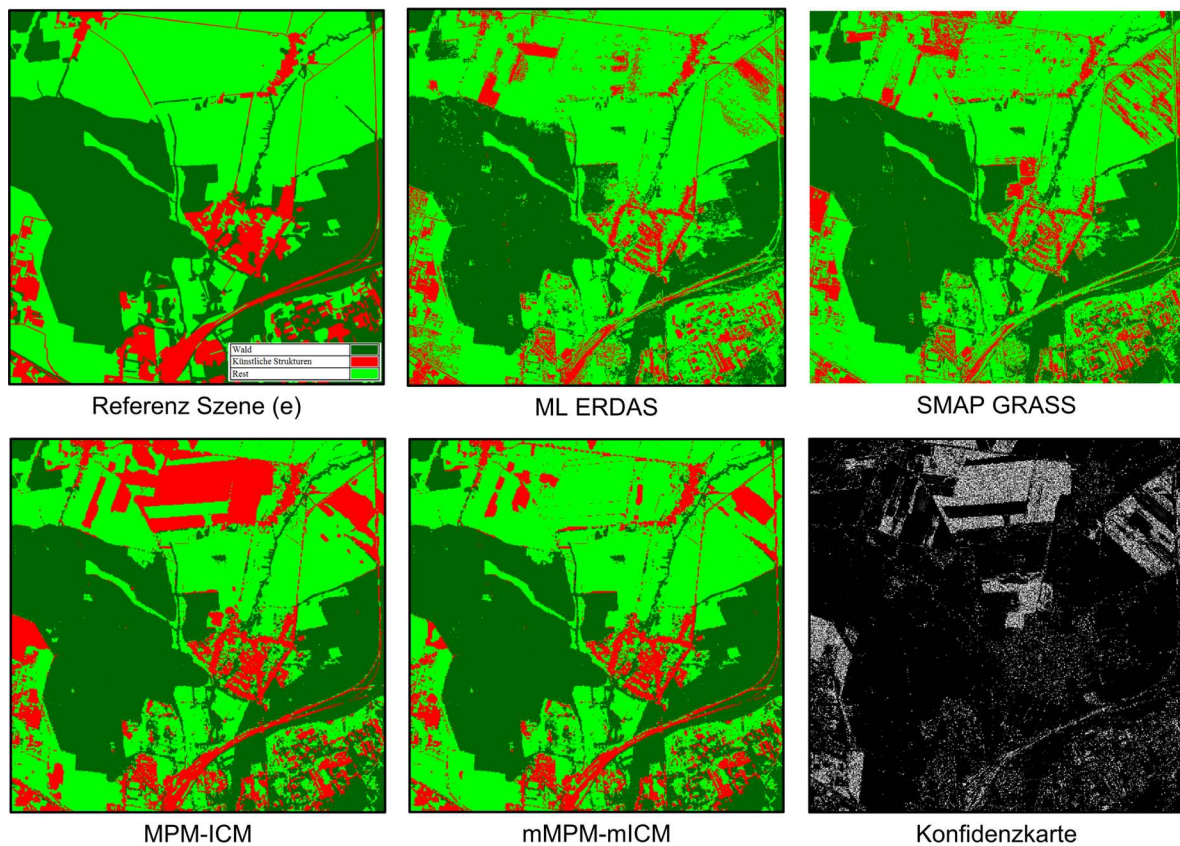


Abbildung 6.21: Konfidenzkarte und Ergebnisse der Methoden ML, SMAP, MPM-ICM, mMPM-mICM für Szene (e).

ter Verwendung der Level (S^N, S^{N-2}) können diese großflächigen Fehlklassifikationen massiv eingedämmt werden. Dieser Effekt ist offensichtlich das Resultat aus der generalisierten Modellierung der Klassen *Wald* und *Rest* in Ebene S^{N-2} einerseits und aus der Vernachlässigung

unsicher klassifizierbarer Bildinformation bei der modifizierten MPM-Inferenz andererseits. Dieses Beispiel zeigt, dass durch eine generalisierte Modellierung in einer höheren Ebene, was in diesem Fall die Vereinigung der Zielklassen *Künstliche Strukturen* und *Rest* in Ebene S^N in die Klasse *Rest* in Level S^{N-2} bedeutet, eine Unterdrückung von großflächigen Fehlklassifikationen in Level S^N mit Hilfe entsprechender Übergangswahrscheinlichkeiten bewirken kann. Aufgrund der markanten Blockstrukturen im Ergebnis der modifizierten MPM-Inferenz ist mit einer Gesamtklassifikationsgüte von 82.36 % zunächst keine große Verbesserung gegenüber der ML-Klassifikation zu erkennen. Dennoch werden hier die großflächigen Fehlklassifikationen im Bereich der Felder (nördlicher Bereich) stark eingeschränkt. Die anschließende Anwendung der modifizierten ICM-Inferenz auf dieses Zwischenergebnis führt wiederum zu einer Auflösung der Blockstrukturen, so dass sich bei dieser Szene ein signifikanter Genauigkeitsgewinn von 7.50 % gegenüber der ursprünglichen MPM-Inferenz ergibt (79.88 % bei MPM-Inferenz und 87.38 % bei mMPM-mICM-Inferenz). Die Konsumentengüte der Klasse *Künstliche Strukturen* aus der mMPM-mICM-Inferenz ist jedoch mit 64.73 % sogar etwas niedriger als mit der ML-Klassifikation (68.28 %), was zeigt, dass auch hier noch sehr viele Fehlklassifikationen erfolgen. Die Konfidenzkarte weist deutlich die großflächigen problematischen Bereiche aus.

6.8.5 Ergebnisse für Szene (f)

Anhand der Ergebnisse von Szene (f) wird ebenfalls der Mehrwert des hybriden Modells deutlich. Laut Tabelle 6.8 unterscheiden sich hier die Gesamtklassifikationsgüten von ML-Klassifikation und MPM-Inferenz um 5.27 % und von ML-Klassifikation und mMPM-mICM-Inferenz um 6.16 %. Bei der ML-Klassifikation sowie der SMAP-Inferenz fallen mit 39.39 % und 28.75 % sehr niedrige Produzentengenauigkeiten der Klasse *Wasser* auf. Mit der Kombination mMPM-mICM kann diese auf 78.98 % verbessert werden, was auch in Abbildung 6.22 gut zu erkennen ist. Generell fällt auf, dass durch jeglichen Kontexteinbezug eine starke Glättung im Stadtbereich bewirkt wird.

6.8.6 Zusammenfassung

Der qualitative und quantitative Vergleich der Ergebnisse in diesem Abschnitt zeigt deutlich den Mehrwert des hybriden Modells gegenüber der ML-Klassifikation, der SMAP-Inferenz sowie der ursprünglichen MPM-Inferenz nach [Laferté et al., 2000]. Das sehr starre und „aufgesetzte“ Quadtree-Modell führt bei jedem Bild zu Segmenten, die Mischklassen aufweisen. Das auf GMMs basierende Datenmodell kann diese Segmente mitunter nicht sinnvoll klassifizieren, was durchaus auch Fehlklassifikationen unter Einbezug von Kontextinformation bewirkt. Treten Konfusionen durch das Datenmodell kleinflächig auf, so können diese teilweise durch die hierarchische MPM-Inferenz und darüber hinaus weiter mittels der ICM-Inferenz verringert werden (Glättung). Bei großflächigen Fehlklassifikationen, wie beispielsweise in Szene (f) bringt dieser Ansatz jedoch mitunter keinen Mehrwert. Die hier dargestellten Beispiele belegen, dass groß- und kleinflächige Fehlklassifikationen sowie Fehlklassifikationen in Bereichen von Klassenübergängen mit Hilfe der Multiskalenprozessierung kombiniert mit modifizierten Inferenzverfahren stark verringert werden.

Die Gesamtklassifikationsgüten der kombinierten mMPM-mICM-Inferenz stellen in drei der sechs Beispiele die besten Ergebnisse dar. Die Unterschiede dieser Methode zu den jeweils besten Ergebnissen der anderen drei Szenen unterscheiden sich maximal um 1.96 % (Szene

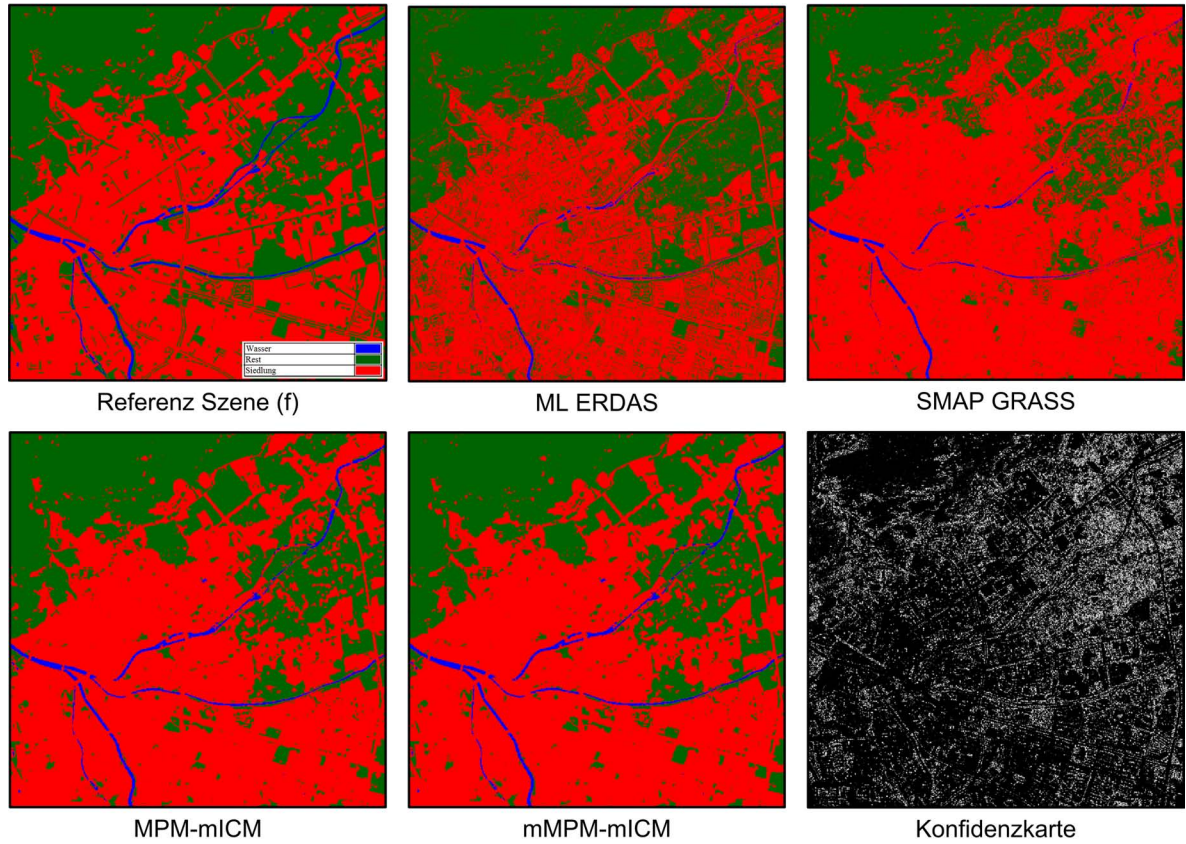


Abbildung 6.22: Konfidenzkarte und Ergebnisse der Methoden ML, SMAP, MPM-mICM, mMPM-mICM für Szene (f).

(d)), so dass eine generelle Anwendung dieser Kombination mit dem hybriden Modell sinnvoll erscheint. Es ist allerdings zu beachten, dass mit jedem Prozessschritt auch zusätzliche Rechenzeit erforderlich wird. Die iterative Schätzung der Parameter des *a priori* Modells ist dabei der aufwändigste Schritt. Besonders in sehr zeitkritischen Fällen ist vom Bildinterpretieren zu entscheiden, wie weit die Prozesskette durchlaufen wird. Die Ergebnisse bezüglich Szene (d) verdeutlichen, dass mit dem vorliegenden Modell nicht in allen Fällen eine Verbesserung gegenüber einer ML-Klassifikation erreicht werden kann. Um kleine Strukturen zu erhalten ist bezüglich der Inferenzverfahren auf eine entsprechende Definition der Parameter α (mMPM) sowie β (mICM) zu achten.

Der Aufwand einer Modellierung von thematischen Klassen in einem Level $S \neq S^N$ sollte nur dann investiert werden, wenn die jeweils definierten thematischen Klassen gut trennbar sind, so dass neue Information in das Level S^N gebracht werden kann. Eine Beurteilung der jeweiligen Datenmodelle kann durch visuelle Begutachtung der Ergebnisse der entsprechenden ML-Klassifikation erfolgen. Für die Bereitstellung der Trainingsdaten für das *a priori* Modell sollte nach Abschnitt 6.6 eine Sicherheitswahrscheinlichkeit von $\alpha = 0.8$ oder $\alpha = 0.9$ gewählt werden.

6.9 Aussagekraft der Konfidenzkarte

In diesem Abschnitt wird untersucht, inwieweit eine signifikante Entropie der *a posteriori* Randdichten tatsächlich ein Indikator für Fehlklassifikationen ist. Dazu werden die jeweiligen Entropiewerte aus der MPM-Schätzung für drei der sechs Szenen mit den tatsächlichen Fehlklassifikationen verglichen. Ziel ist die Ableitung eines Grenzwerts für die Entropie, bei dem die Fehlklassifikationen bestmöglich erfasst werden, so dass der modifizierte ICM-Algorithmus einerseits möglichst viele falsch klassifizierte Segmente einbezieht und andererseits signifikant Rechenzeit einspart. Für drei Szenen werden im Folgenden bei unterschiedlichen Schwellwerten T_H der Entropie H der Zusammenhang zwischen einer Entropie $H > T_H$ und den tatsächlichen Fehlklassifikationen aus der jeweils besten MPM-Inferenz aus Abschnitt 6.7 untersucht. In Tabelle 6.10 ist je Szene die prozentuale Gesamtfehlerrate ER der MPM-Inferenz angegeben. In den Zeilen der Tabelle sind des Weiteren in Abhängigkeit des Grenzwerts T_H jeweils die Anteile der Segmente $\{x\}_{T_H < H_i}$, deren Entropiewert diesen Grenzwert überschreiten sowie die Anteile $\{x\}_{T_H < H,e}$ der darin enthaltenen Fehlklassifikationen jeweils in % von der Gesamtmenge aller Segmente dargestellt. Die grau hinterlegte Spalte kennzeichnet den bei

T_H	Szene und MPM-Fehlerrate ER in %					
	(a), ER: 3.24 %		(e), ER: 20.76 %		(f), ER: 10.86 %	
	$\{x\}_{T_H < H}$	$\{x\}_{T_H < H,e}$	$\{x\}_{T_H < H}$	$\{x\}_{T_H < H,e}$	$\{x\}_{T_H < H}$	$\{x\}_{T_H < H,e}$
1.0	0.0	0.0	0.02	0.01	0.00	0.00
0.5	0.44	0.23	2.48	1.20	4.60	2.04
0.1	0.99	0.46	5.95	2.87	9.84	3.92
0.05	1.22	0.54	7.74	3.97	12.03	4.60
0.01	1.80	0.71	14.16	8.80	17.28	6.07
0.005	2.05	0.77	15.27	9.21	19.68	6.58
0.001	2.73	0.93	17.31	9.88	25.60	7.51
0.0005	3.07	0.99	18.16	10.15	28.50	7.82
0.0001	4.05	1.12	20.03	10.85	36.32	8.51
0.00005	4.60	1.19	20.85	11.16	40.27	8.78
0.00001	6.06	1.32	22.61	11.83	49.73	9.32

Tabelle 6.10: Vergleich der Fehlerraten ER der MPM-Inferenz mit den prozentualen Anteilen der Segmente, die eine Entropie $H > T_H$ aufweisen und die zudem falsch klassifiziert wurden.

dem Methodenvergleich im vorherigen Abschnitt angesetzten Schwellwert $T_H = 0.0001$. Bei diesem Wert sind demnach in der daraus resultierenden Menge der Segmente für Szene (a) 34.57 %, für Szene (e) 52.26 % und für Szene (f) 78.36 % aller falsch klassifizierten Segmente enthalten, wobei die Menge der Segmente $\{x\}_{T_H < H}$ bei Szene (a) lediglich 4.5 %, für Szene (e) 20.03 % und für Szene (f) sogar 36.32 % aller Segmente enthält. Eine allgemeine Aussage über einen zu wählenden Schwellwert T_H lässt sich daran nicht ableiten, jedoch wird damit gezeigt, dass mit $T_H = 0.0001$ bei diesen drei Szenen mindestens 1/3 aller falsch klassifizierten Segmente erfasst werden und damit tatsächlich ein Zusammenhang zwischen der Entropie und den Fehlklassifikationen besteht.

6.10 Zusammenfassung

Ziel der Evaluierungen in diesem Abschnitt war, die einzelnen Schritte für das Training des hybriden Modells bezüglich ihrer Robustheit zu untersuchen sowie die Art und Ausprägung der Abhängigkeiten der Prozessierungsergebnisse von den interaktiv zu definierenden Parametern und Trainingsdaten herauszustellen. Des Weiteren wurde der Mehrwert der Multiskalenprozessierung untersucht und die Methode respektive verschiedene Prozessierungsergebnisse dieser mit zwei weiteren Klassifikationsverfahren verglichen. Die Untersuchungen wurden mit insgesamt sechs verschiedenen Subszenen des Sensors IKONOS durchgeführt. Die Ergebnisse der Experimente können wie folgt zusammengefasst werden:

- **Datenmodell:** Anhand der Kreuzvalidierung in Abschnitt 6.4 hat sich gezeigt, dass die Güte eines geschätzten Datenmodells maßgeblich von der Qualität der Trainingsdaten und weniger von der Menge der Trainingsdaten und Merkmale abhängt. Vor allem bei spektral inhomogenen und schwer trennbaren Klassen ist auf eine sorgfältige und repräsentative Wahl der Trainingssegmente zu achten. Dies korrespondiert mit den allgemein bekannten Anforderungen bei der Verwendung von GMMs. Um die Rechenzeit möglichst gering zu halten, wurden bei den Untersuchungen nicht mehr als zehn Merkmale verwendet.
- **A priori Modell:** Durch die Untersuchung in Abschnitt 6.5 konnte gezeigt werden, dass die Schätzung der Parameter des *a priori* Modells nur bedingt von der Initialisierung dieser sowie der Menge der Trainingsdaten abhängt. Wie viele und welche Trainingssegmente bei der Schätzung herangezogen werden, hängt von dem bei der Vorklassifikation zu definierenden Fehlerniveau $\alpha \in [0, 1]$ ab. Bei einem wachsenden Wert für α ist zu erwarten, dass immer weniger Segmente mit der korrespondierenden Sicherheitswahrscheinlichkeit $P = 1 - \alpha$ eindeutig einer Klasse zugewiesen werden können.

Werden mehr als 40 % aller Bildsegmente je Klasse für das Training herangezogen, so ergeben sich nur sehr geringe Unterschiede in den Schätzergebnissen, auch wenn unterschiedliche Initialisierungen angesetzt werden. Werden weniger als 40 % der Daten für das Training herangezogen, so macht sich die bei der EM-Schätzung angenommene Gleichverteilung für die Segmente, die keine Klassenzugehörigkeit haben, in den geschätzten Parametern und somit auch in den Inferenzergebnissen bemerkbar. In diesem Fall werden die Übergangswahrscheinlichkeiten für gleiche Klassen im Schätzergebnis kleiner, was eine Abschwächung des Glättungseffekts bei der Inferenz bewirkt. Bei Bildern mit kleinen Strukturen, wie in den Szenen (f) und (g) kann sich dies wiederum positiv auswirken, da diese nun weniger vernachlässigt werden. Bei Bildinhalten mit größeren homogenen thematischen Bildbereichen kann sich dies jedoch auch negativ auswirken (Szene (b) und (f)). Insgesamt bewegen sich die Unterschiede der Gesamtklassifikationsgüten in den Inferenzergebnissen bei den unterschiedlichen Konfigurationen jedoch in einem Bereich von ca. 0.5 %. Eine Initialisierung mit gleichverteilten Übergangswahrscheinlichkeiten ergab keine sinnvollen Ergebnisse und sollte deshalb vermieden werden. Wird eine Initialisierung ungleich der Gleichverteilung verwendet, so ergeben sich mit der MPM-Inferenz stets bessere Ergebnisse, als mit der ML-Klassifikation. Die Gesamtklassifikationsgüten verbessern sich durch den Einbezug von hierarchischem Kontext bei den vorliegenden Szenen gegenüber einer ML-Klassifikation mit dem selben Datenmodell lediglich um maximal 1.4 % (Szene (b)).

- **Bereitstellung von Trainingsdaten für das *a priori* Modell:** In diesem Experiment wurde geprüft, wie sich die Ergebnisse der EM-Schätzung der Übergangswahrscheinlichkeiten verhalten, wenn die Trainingsdaten dafür nicht zufällig aus einer Referenz, sondern mit Hilfe des Datenmodells bereitgestellt werden. Dazu wurden durch Variation des Fehlerniveaus α unterschiedliche Trainingsdatensätze je Bild mittels beschränkter ML-Klassifikation generiert und die erreichten Gesamtklassifikationsgüten der entsprechenden MPM-Inferenzergebnisse gegenübergestellt. Auch wenn mit einem sehr hohen Fehlerniveau von $\alpha = 0.9$ nur sehr wenige Trainingsdaten bereitgestellt werden, ergaben sich dabei für fast alle Szenen die günstigsten *a priori* Modellparameter respektive besten Inferenzergebnisse. Diese unterscheiden sich nur sehr gering von den Ergebnissen, die mit zufällig aus der Referenz generierten Daten erreicht wurden.
- **Multiskalenprozessierung:** Die Untersuchungen bezüglich des *a priori* Modells ergaben, dass durch den Einbezug von hierarchischem Kontext für die hier prozessierten Szenen die Gesamtklassifikationsgüte um maximal 1.4 % gegenüber der ML-Klassifikation unter Verwendung des gleichen Datenmodells verbessert werden konnte. In Abschnitt 6.7 wurde untersucht, inwieweit die Modellierung von thematischen Klassen in mehreren Hierarchieebenen einen Mehrwert respektive eine Verbesserung in den Klassifikationsergebnissen der MPM-Inferenz bewirken kann. Die besten Ergebnisse konnten unter Verwendung von Datenmodellen auf zwei beziehungsweise drei Hierarchieebenen erreicht werden. Die erreichten Verbesserungen der Gesamtklassifikationsgüten liegen in einem Bereich von bis zu 2% gegenüber einer ML-Klassifikation, jedoch zeigt eine Gegenüberstellung der Konsumenten- und Produzentengenauigkeiten der einzelnen Klassen sowie die visuelle Gegenüberstellung der Klassifikationsergebnisse einen deutlichen Mehrwert des hierarchischen Modells. Die Multiskalen-Prozessierung bewirkt eine sichtliche Glättung homogener Flächen sowie größtenteils Verbesserungen der Konsumenten- und Produzentengenauigkeiten gegenüber einer ML-Klassifikation. Fehlklassifikationen von Mischsegmenten treten hier jedoch dennoch auf.
- **Methodenvergleich:** Bei dem in Abschnitt 6.8 durchgeführten Vergleich des hybriden Modells mit zwei weiteren gängigen Methoden wurden die jeweils besten Konfigurationen von Datenmodellen aus den Untersuchungen der Multiskalenprozessierung herangezogen. Das hybride Modell wurde mit einer pixelbasierten ML-Klassifikation sowie der hierarchischen SMAP-Inferenz verglichen. Die entsprechenden Ergebnisse dieser Klassifikationen wurden unterschiedlichen Ergebnissen des hybriden Modells gegenüber gestellt. Es konnte ein deutlicher Mehrwert des hybriden Modells gegenüber der ML-Klassifikation, der SMAP-Inferenz sowie der ursprünglichen MPM-Inferenz nach [Laferté et al., 2000] herausgestellt werden. Probleme bei der spektralen Trennbarkeit von thematischen Klassen mit dem Datenmodell können teilweise durch die hierarchische MPM-Inferenz (Glättung) und darüber hinaus weiter mittels der ICM-Inferenz verringert werden, sofern die resultierenden Fehlklassifikationen kleinflächig auftreten. Die hier dargestellten Beispiele belegen, dass groß- und kleinflächige Fehlklassifikationen sowie Fehlklassifikationen in Bereichen von Klassenübergängen mit Hilfe der Multiskalenprozessierung kombiniert mit modifizierten Inferenzverfahren stark verringert werden können. Auf der Grundlage der Ergebnisse wird eine generelle Anwendung der Kombination der hierarchischen mMPM-Inferenz und der planaren mICM-Inferenz empfohlen.

- **Aussagekraft der Konfidenzkarte:** In Abschnitt 6.9 konnte anhand von drei der sechs Szenen gezeigt werden, dass die Konfidenzkarte zu einem gewissen Maße tatsächlich ein Indikator für Fehlklassifikationen der MPM-Inferenz ist. Bei dem Methodenvergleich in Abschnitt 6.8 wurde der Schwellwert für die Entropie mit $T_H = 0.0001$ definiert. Hier konnte gezeigt werden, dass damit in der daraus resultierenden Segmentmenge große Anteile der falsch klassifizierten Segmente enthalten sind. Ein alleiniger Einbezug dieser Segmente bei der ICM-Inferenz führt zu einer signifikanten Verringerung der Laufzeit, wobei quantitativ vergleichbare Ergebnisse, wie mit dem ursprünglichen ICM-Algorithmus, erreicht werden. Während der ICM-Algorithmus für Szene (a) 189 sec benötigt (Gesamtklassifikationsgüte: 97.31 %), hat die modifizierte Version das Abbruchkriterium in lediglich 8 sec erfüllt (Gesamtklassifikationsgüte: 97.12 %).

Eine detaillierte Darstellung der Laufzeiten der einzelnen Prozessierungsschritte erfolgt aufgrund der nicht optimierten Implementierung nicht. In [Kersten und Gähler, 2010] wird jedoch ein tabellarischer Überblick für einige Teilprozesse gegeben.

7 Zusammenfassung und Ausblick

In dieser Arbeit wurde die Klassifikation von hochauflösenden optischen Satellitenbildern im Kontext von Notfall- und Krisenkartierungen behandelt. Der Bedarf an neuen Methoden zur Bildverarbeitung in diesem Bereich ist durch verschiedene Gründe evident. Zum ersten ist in den vergangenen Jahren die Verfügbarkeit von hochauflösenden und sehr hochauflösenden Satelliten- und Luftbildern für zivile Anwendungen stark gestiegen. Eine rechnergestützte semiautomatische oder gar vollautomatische Prozessierung (Klassifikation von Landbedeckung, Landnutzung sowie Informationsextraktion) dieser Bilder ist aufgrund ihres sehr hohen Informationsgehalts sehr komplex. Zum zweiten bestehen für einen operationellen Notfall- und Krisenkartierungsdienst klar definierte Rahmenbedingungen und Anforderungen bezüglich der bereitgestellten Informationsprodukte. So wird vor allem in der *emergency phase*, unmittelbar nach Eintreten eines Notfalls, ein zeitnah geliefertes, jedoch bezüglich der räumlichen und thematischen Genauigkeit „etwas ungenaueres“ Informationsprodukt einem sehr genauen und zeitaufwändig generierten Produkt vorgezogen. Auch die Verfügbarkeit von Zusatzdaten, wie beispielsweise digitale Höhenmodelle, Vektordaten oder zusätzliche Satellitenbilder vom betroffenen Gebiet ist in diesem Zeitrahmen oft sehr beschränkt. Des Weiteren spielt auch die Verlässlichkeit der angewendeten Methoden zur Bildverarbeitung eine sehr wichtige Rolle. Diese Aspekte führen dazu, dass bei Notfall- und Krisenkartierungen häufig semiautomatische Methoden, manuelle Zwischen- und Nachbearbeitungsschritte oder gar vollständig manuelle Verfahren (visuelle Interpretation) zur Interpretation hochauflösender Satellitenbilder angewendet werden.

Entgegen dem allgemeinen Trend hin zur vollautomatischen Bildprozessierung wurde aufgrund der oben angeführten Gründe in dieser Arbeit ein interaktives Rahmenwerk zur raschen Klassifikation und Vorklassifikation hochauflösender optischer Satellitenbilder konzipiert und evaluiert. Dieses Rahmenwerk hat nicht den Anspruch, die sehr komplexe Aufgabe der Extraktion von höherwertigen und pixelgenauen Informationen (beispielsweise Objekt- oder Lineamentextraktion) aus Satellitenbildern zu leisten, sondern stellt vielmehr ein Konzept für einen allgemeinen und übertragbaren Ansatz zur Klassifikation von Landbedeckung in wenige thematische Klassen bei den oben genannten Rahmenbedingungen dar. Die verwendeten Methoden wurden auf der Grundlage theoretischer Überlegungen gewählt, wobei insbesondere die schnelle Laufzeit entscheidend war. Daneben bestand die Forderung, dass die Prozessierung lediglich auf dem Satellitenbild selbst sowie den interaktiven Eingaben des Bildinterpreteten beruhen soll (eingeschränkte Datengrundlage). Die geforderte Übertragbarkeit der Methode bezieht sich sowohl auf Bildprodukte verschiedener Sensoren sowie auf unterschiedlichste Bildinhalte und thematische Fragestellungen.

In Abschnitt 1 erfolgte, basierend auf den definierten Zielen und den herrschenden Rahmenbedingungen in dieser Arbeit, eine erste Eingrenzung der Methodik auf einen segmentbasierten Ansatz, bei dem aus Gründen der Laufzeit sowie der obigen Zielsetzung einfache quadratische Bildelemente verwendet werden sollen. Des Weiteren wurde der Einbezug von Kontextinfor-

mation für die Auflösung beziehungsweise Verringerung von Mehrdeutigkeiten aufgrund spektraler Ähnlichkeiten von thematischen Klassen vorgeschlagen und diesbezüglich in Abschnitt 2 planare und hierarchische Ansätze zu graphenbasierten Bildmodellierung behandelt. Aufgrund von Unsicherheiten, Rauschen und spektralen Mehrdeutigkeiten existiert für Klassifikationsprobleme üblicherweise keine exakte Lösung. Ansätze der statistischen graphenbasierten Bildmodellierung beruhen auf den Prinzipien der statistischen Optimierung und können in-exakte jedoch optimale Lösungen für verschiedenste Problemstellungen finden. Weitere große Vorteile gegenüber anderen Verfahren sind die ebenfalls auf Optimierung basierenden Methoden der Schätzung der Modellparameter sowie die Möglichkeit der Multiskalenprozessierung. Daher erfolgt die Anwendung eines solchen Ansatzes in der vorliegenden Arbeit. Es ist dabei zu beachten, dass die Qualität der Ergebnisse hier stark von der Wahl eines adäquaten Modells abhängt.

In Abschnitt 2.2 wurde ein umfassender Überblick über existierende planare und hierarchische Methoden der graphenbasierten Bildmodellierung gegeben, wobei die jeweilige Anwendung sowie die grundlegenden Eigenschaften und Unterschiede herausgestellt wurden. Basierend auf dieser Gegenüberstellung wurde die Anwendung eines Ansatzes der hierarchischen graphenbasierten Bildmodellierung in dieser Arbeit gewählt. Gründe für diese Wahl sind die damit einhergehenden attraktiven Eigenschaften der Modellierung weitreichender kontextueller Relationen sowie der Möglichkeit der Multiskalenprozessierung. Des Weiteren ist bei einer Wahl eines Graphen in Form einer Baumstruktur immer die Möglichkeit der exakten und nicht iterativen Inferenz gegeben, was für Anwendungen im Bereich Notfall- und Krisenkartierungen sehr attraktiv ist. Aufgrund der Verwendung quadratischer Bildelemente liegt somit ein Quadtree-Bildmodell nahe.

In Abschnitt 3 wurden die in der Literatur vorgestellten Ansätze der graphenbasierten Bildmodellierung genauer untersucht. Im Zuge dessen wurden die Verfahren in die Untergruppen kausale und nicht kausale Modelle eingeteilt, wobei jeweils sowohl hierarchische als auch planare Modelle behandelt wurden. Für jeden dieser vier Ansatztypen wurden die statistischen Eigenschaften sowie Methoden der Parameterschätzung und Inferenz dargestellt. Neben einer detaillierten Darstellung der Methoden erfolge in Abschnitt 3.5 eine theoretische Gegenüberstellung generativer und diskriminativer Verfahren. Während bei generativen Verfahren die gemeinsame Verteilung $P(x, y) = p(y|x)P(x)$ der gesuchten Klassenlabels x sowie der entsprechenden Bildmerkmale y modelliert wird, wobei die Zufallsvariablen X als MRF angesehen werden und $p(y|x)$ explizit durch ein Datenmodell beschrieben wird, erfolgt in diskriminativen Ansätzen eine direkte Modellierung der *a posteriori* Verteilung $P(x|y)$. Daraus resultiert, dass eine explizite Formulierung eines Datenmodells vermieden wird und somit keine Annahmen beispielsweise über die Verteilung der Daten getroffen werden müssen. Dies erlaubt eine Verwendung von überlappenden Bildmerkmalen. Des Weiteren können die kontextuellen Bedingungen auch in Abhängigkeit der Bildinformation modelliert werden. Diesen üblicherweise in der Literatur angeführten Vorteilen steht die geforderte Übertragbarkeit der Methode auf verschiedenste thematische Problemstellungen in dieser Arbeit gegenüber. Die Gründe für die Wahl eines generativen Verfahrens in dieser Arbeit sind mit der erforderlichen Merkmalsselektion, der Verfügbarkeit von nur sehr wenigen Trainingsdaten sowie die Schwierigkeit einer allgemein gültigen und sinnvollen Formulierung von überlappenden Bildmerkmalen und kontextuellen Relationen in Abhängigkeit der Bilddaten gegeben.

In Abschnitt 4 wurden der vorgeschlagene generative hybride Ansatz der graphenbasier-

ten Bildmodellierung, die statistischen Eigenschaften dieses sowie die Methoden der Parameterschätzung und Inferenz detailliert behandelt. Das angewendete hierarchische Bildmodell aus [Laferté et al., 2000] und [Pérez et al., 2000] hat neben den oben genannten Vorteilen zwei entscheidende Nachteile. Zum einen führt das starre Quadtree-Bildmodell mitunter zu Blockstrukturen im Inferenzergebnis. Der Einbezug von räumlichem Kontext mittels eines planaren MRFs in der feinsten Quadtree-Ebene kann diesem Effekt entgegenwirken und das hierarchische Inferenzergebnis verbessern. Um den gewinnbringenden Effekt der nicht iterativen und exakten hierarchischen Inferenz nicht zu verlieren, wird ein nachgeschalteter Einbezug von räumlichem Kontext vorgeschlagen, woraufhin hier von einem hybriden Modell gesprochen werden kann. Zum anderen ist durch die quadratischen Bildelemente das Auftreten von Segmenten mit Mischklassen sehr wahrscheinlich, was nach eigenen Erfahrungen bei der hierarchischen Inferenz, aufgrund von Fehlentscheidungen des Datenmodells, leicht zu Fehlklassifikationen führen kann. Daher wird in dieser Arbeit eine modifizierte hierarchische MPM-Inferenz vorgeschlagen. Hier werden Bildinformationen von Segmenten, welche aufgrund des Datenmodells nur unsicher und nicht eindeutig einer Klasse zugeordnet werden können, bei der Inferenz vernachlässigt.

In Abschnitt 5 erfolgte die Integration des hybriden Modells in ein interaktives Rahmenwerk. Dabei wurde die komplette sequentielle Prozesskette sowie alle Teilprozesse detailliert und aus der Sicht eines Anwenders des Rahmenwerks beschrieben. Insbesondere wurden hier die nötigen manuellen Eingaben sowie die verschiedenen Möglichkeiten der Prozessierung mit dem Rahmenwerk herausgestellt. Anhand eines exemplarischen Anwendungsbeispiels wurden zusätzlich die einzelnen Schritte und die Möglichkeiten der Visualisierung von Zwischen- und Endergebnissen mit der vorliegenden prototypischen Implementierung demonstriert.

In Abschnitt 6 wurde das Rahmenwerk konkret auf sechs verschiedene IKONOS Subszenen angewendet, empirisch evaluiert sowie praktisch mit zwei weiteren Klassifikationsmethoden verglichen. Die dafür angewendeten Bildmerkmale sowie die Methode zur Selektion der relevanten Merkmale wurden in den Abschnitten 6.2 und 6.3 behandelt. Anschließend folgte die Beschreibung und Auswertung verschiedener Experimente, deren Versuchsaufbau und Ergebnisse wie folgt zusammengefasst werden können:

- **Untersuchung der Parameterschätzung für das Datenmodell** bezüglich der Abhängigkeit der Modellgüte von der Wahl und Kardinalität der Trainingsdaten einerseits sowie von der Anzahl der Merkmalsraumdimensionen andererseits (Abschnitt 6.4):

Dieses Experiment hat gezeigt, dass die Güte der geschätzten Datenmodelle vergleichsweise gering durch die Anzahl der Trainingssegmente oder der Merkmale, jedoch maßgeblich durch die Wahl der Trainingsdaten beeinflusst wird. Das Maß für die Güte in diesem und den anderen Experimenten ist die Gesamtklassifikationsgüte.

- **Untersuchung der Parameterschätzung für das a priori Modell** bezüglich der Abhängigkeit der Modellgüte wiederum von der Kardinalität der Trainingsdaten sowie von der Initialisierung der Modellparameter (Abschnitt 6.5):

Es konnte für die sechs Szenen gezeigt werden, dass bei unterschiedlichen Initialisierungen ungleich einer Gleichverteilung und einem Anteil der Trainingssegmente von $\geq 40\%$ von der Gesamtmenge der Segmente kaum Unterschiede bei den geschätzten Übergangswahrscheinlichkeiten und somit bei den entsprechenden Inferenzergebnissen

auftreten. Für die Segmente, die nicht in die EM-Schätzung einfließen, wird eine gleiche Wahrscheinlichkeit für das Auftreten aller Klassen angenommen. Werden Anteile an Trainingsdaten $< 40\%$ verwendet, wirkt sich diese Gleichverteilung in Form von leicht abgeschwächten Übergangswahrscheinlichkeiten für gleiche Klassen aus. Dies schwächt wiederum den Glättungseffekt bei der Inferenz ab. Während sich dies bei Bildern mit flächenhaften homogenen Bildbereichen negativ auswirken kann, bleiben feine Strukturen, wie beispielsweise Häuser oder Straßen, eher erhalten. Eine EM-Parameterschätzung mit einer gleichverteilten Initialisierung der Übergangswahrscheinlichkeiten ergab keine sinnvollen Ergebnisse und sollte demnach vermieden werden.

- **Untersuchung der Bereitstellung von Trainingsdaten für das *a priori* Modell** (Abschnitt 6.6):

Hier wurden die Auswirkungen der Wahl des Fehlniveaus α bei der Bereitstellung der Trainingsdaten für die Schätzung der Parameter des *a priori* Modells untersucht. Bei einem Fehlniveau von $0.8 \leq \alpha \leq 0.9$ für den $\chi^2_{f,\alpha}$ -Test der beschränkten ML-Klassifikation konnte für alle sechs Szenen das jeweils beste Inferenzergebnis mit den resultierenden Modellparametern verzeichnet werden. Diese Ergebnisse unterschieden sich bei fünf der sechs Szenen kaum vom besten Ergebnis aus Abschnitt 6.5.

- **Untersuchung hinsichtlich des Mehrwerts der Multiskalenprozessierung mit dem hierarchischen Modell** (Abschnitt 6.7):

Der Mehrwert der Multiskalenprozessierung wurde hier durch die Kombination von Datenmodellen auf drei verschiedenen Hierarchieebenen untersucht, wobei jeweils die Ergebnisse der entsprechenden MPM-Inferenz sowie einer ML-Klassifikation gegenüber gestellt wurden. Es konnte in allen Fällen eine leichte Verbesserung in den Gesamtklassifikationsgüten sowie hinsichtlich der Konfusionen zwischen den thematischen Klassen beobachtet werden. Insgesamt wurde festgestellt, dass unter der Verwendung des Quadtree-Bildmodells mit der hierarchischen Inferenz überwiegend durch Glättungseffekte leichte Verbesserungen in den Klassifikationsgüten gegenüber einer ML-Klassifikation erreicht werden können. Großflächige Fehlklassifikationen sowie Fehlklassifikationen in Übergangsbereichen thematischer Klassen, wie sie auch unter alleiniger Verwendung des Datenmodells (ML-Klassifikation) auftreten, können jedoch offensichtlich mit diesem Modell nicht ausreichend verringert werden.

- **Untersuchung der Aussagekraft der Konfidenzkarte** (Abschnitt 6.9): Hier wurde anhand von drei Szenen gezeigt, dass die aus den Entropien der *a posteriori* Randdichten generierte Konfidenzkarte tatsächlich zu einem gewissen Maß ein Indikator für falsch klassifizierte Bildsegmente darstellt.

In Abschnitt 6.8 wurde das hybride Modell einschließlich der modifizierten Inferenzmethoden für das planare und das hierarchische Modell anhand der sechs Szenen qualitativ und quantitativ mit einer ML-Klassifikation in ERDAS sowie der hierarchischen SMAP-Inferenz in GRASS verglichen. Vor allem durch die Anwendung der modifizierten Inferenzmethoden konnte für fünf der sechs Szenen ein teilweise sehr deutlicher Mehrwert des hybriden Modells gegenüber der Methoden ML und SMAP sowie der ursprünglichen hierarchischen Inferenz festgestellt werden. Die hier dargestellten Beispiele belegen, dass Mehrdeutigkeiten zwischen

thematischen Klassen, sowohl klein- als auch großflächige Fehlklassifikationen sowie Fehlklassifikationen in Bereichen von Klassenübergängen mit Hilfe der Multiskalenprozessierung, kombiniert mit den modifizierten Inferenzverfahren mMPM und mICM, signifikant verringert werden können.

Die entscheidenden Vorteile der Methode gegenüber anderen derzeit operationell eingesetzten Verfahren sind mit der festen und sequentiellen Vorgehensweise bei der Prozessierung, die Kontrollen nach jedem Zwischenschritt erlaubt, der überschaubaren Anzahl manuell zu definierender Parameter sowie der übertragbaren Anwendbarkeit gegeben. Durch die graphenbasierte Bildmodellierung wird darüber hinaus der Einbezug von hierarchischem und räumlichem Kontext sowie eine Multiskalenprozessierung ermöglicht, wobei die auf statistischer Optimierung beruhende Parameterschätzung und Inferenz für jedes Klassifikationsproblem individuell erfolgt.

Wenngleich die Ergebnisse mitunter einen Mehrwert gegenüber anderen Methoden belegen, ist eine Anwendung des Rahmenwerks in einem operationellen Kartierungsdienst in Abhängigkeit der jeweiligen thematischen Fragestellung zu beurteilen. Das Rahmenwerk ist mit dem derzeitigen Stand als Konzept für eine allgemeine Herangehensweise der Klassifikation hochauflösender optischer Satellitenbilder unter den definierten Rahmenbedingungen zu sehen und lässt somit bisher spezielle krisenrelevante thematische Fragestellungen außen vor.

Eine sehr aktuelle und relevante thematische Problemstellung ist die Detektion von Wasser in Satellitenbildern. Üblicherweise gehen Flutereignisse mit schlechten Wetterlagen einher, so dass sich die Verwendung von SAR-Daten für die Erfassung von Wasserflächen aufgrund ihrer weitgehenden Wetterunabhängigkeit eignet. Glatte Wasseroberflächen können hier aufgrund der seitlichen Aufnahmegeometrie in der Regel gut vom Rest getrennt werden. Unterflutete Vegetation wiederum erscheint mitunter sehr hell im Amplitudenbild (*double-bounce* Effekt), so dass die Trennung von anderen hellen Bildbereichen sehr schwierig sein kann. Hinzu kommt die Schwierigkeit der Klassifikation bewegter und rauer Wasseroberflächen in SAR-Daten. Vor allem vor dem Hintergrund der vielfältigen Bildmerkmale sowie der Multiskalenprozessierung ist daher die Anwendung des Rahmenwerks für die Klassifikation von Wasser ein wichtiger Aspekt für weiterführende Untersuchungen. Auf der Grundlage von Markov-Zufallsfeldern könnte, neben der Multiskalenprozessierung, auch die Fusion, beispielsweise von optischen und SAR-Daten, untersucht werden.

Für verschiedenste thematische Fragestellungen erscheint eine Untersuchung bezüglich der Eingrenzung der relevanten Bildmerkmale in unterschiedlichen Skalenbereichen sehr gewinnbringend, denn der Aufwand der Selektion relevanter Merkmale würde sich daraufhin stark verringern. Insbesondere vor dem Hintergrund der Klassifikation von anthropogen beeinflussten Bildbereichen sollte des Weiteren eine Erweiterung des Rahmenwerks um weitere Strukturmerkmale, beispielsweise unter Verwendung von Wavelets sowie Gradienteninformation, erfolgen.

Auch wenn die Laufzeit ein entscheidendes Kriterium bei der Auswahl der verschiedenen Methoden war, erfolgte in dieser Arbeit keine Untersuchung dieser, was mit der vorliegenden suboptimalen Implementierung begründet werden kann. Die vollständige Prozessierung einer Szene der Größe 1024×1024 Pixel beläuft sich für einen Benutzer, der Erfahrungen mit dem Rahmenwerk hat, derzeit auf etwa eine Stunde. Vor dem Hintergrund einer möglicherweise operationellen Anwendung ist daher eine generelle Optimierung der Implementierung ein wichtiger Punkt für zukünftige Arbeiten, so dass schnellere Rechenzeiten einerseits und die Prozessierung großer Bilder andererseits erfolgen kann. Eine exemplarische Darstellung der Laufzeiten

für die verschiedenen Schritte mit der vorliegenden Implementierung wird in [Kersten und Gähler, 2010] gegeben. Aus dieser wird deutlich, dass vor allem die Merkmalsselektion sehr rechenintensiv ist. Daher sollte, vor allem vor dem Hintergrund einer möglichen Erweiterung der Methode um weitere Bildmerkmale, die Verwendung einer alternativen Merkmalsselektion in Erwägung gezogen werden. Relevante Ansätze dafür sind mit der simultanen Bayes'schen Merkmals- und Modellselektion für GMMs [Constantinopoulos et al., 2006] sowie der simultanen lokalen Bayes'schen Merkmals- und Modellselektion [Li et al., 2009] gegeben.

Für bestimmte thematische Fragestellungen ist es denkbar, dass die explizite Erhaltung von feinen Strukturen, beispielsweise linienhaften Elementen, gewünscht ist. Vor diesem Hintergrund sollte die Verwendung von adaptiven Nachbarschaften, wie beispielsweise in [Zhong et al., 2007], untersucht werden.

Die konzipierte Methode wurde im Rahmen dieser Arbeit lediglich auf verschiedene Szenen des Sensors IKONOS angewendet. Es sind jedoch weitere Untersuchungen, wie beispielsweise die Anwendung auf QuickBird-Szenen in [Kersten et al., 2010], nötig, um die Übertragbarkeit der Methode weiter zu untersuchen und daraufhin Weiterentwicklungen und Anpassungen vornehmen zu können.

Bei den Experimenten in dieser Arbeit wurde zudem der Aspekt der generalisierten Modellierung von thematischen Klassen in größeren Hierarchieebenen nicht ausreichend detailliert behandelt. In vielen Ansätzen der hierarchischen Bildmodellierung kann die Bedeutung eines Knotens in einer groben Hierarchieebene nicht semantisch interpretiert werden. Bei Fernerkundungsdaten können die thematische Klassen in feinen Ebenen jedoch zu generalisierten Klassen in höheren Ebenen zusammengesetzt werden, womit allen Bildsegmenten eine semantische Bedeutung zukommt. Dieser Gedanke wird in [Kolaczyk et al., 2005] aufgegriffen und sollte auch für dieses Rahmenwerk detaillierter beleuchtet werden.

Der praktische Vergleich der Methode erfolgte hier mit einem weiteren hierarchischen generativen Verfahren, der hierarchischen SMAP-Inferenz. Ein entsprechender Vergleich zwischen hierarchischen generativen und diskriminativen Verfahren ist nach Wissen des Autors bisher in der Literatur noch nicht erfolgt und ist daher ein wichtiger und interessanter Aspekt für zukünftige Untersuchungen. Insbesondere wird momentan an einem Vergleich mit dem TCRF-Ansatz nach [Awasthi et al., 2007] gearbeitet.

Der offensichtliche Nachteil der verwendeten Quadtree-Struktur ist das damit einhergehende Auftreten von Segmenten mit Mischklassen. Der Einfluss des Datenmodells wird bei der modifizierten hierarchischen Inferenz durch einen harten Schwellwert gesteuert, woraufhin dieses entweder völlig ignoriert oder in Gänze berücksichtigt wird. Eine weichere Regulierung des Einflusses des Datenmodells könnte beispielsweise unter Verwendung von *fuzzy* Kurven erfolgen.

Fehlklassifikationen aufgrund von Segmenten mit Mischklassen könnte mit der Anwendung einer Bildsegmentierung, wie beispielsweise in [He et al., 2006] und [Martinis et al., 2011], entgegengewirkt werden. Eine sehr schnelle Bereitstellung von Segmenten in der feinsten Hierarchieebene kann beispielsweise mit dem optimierten Wasserscheiden-Verfahren nach [De Smet und Pries, 2000] erreicht werden, wobei für die Multiskalenprozessierung zusätzlich die Aggregation von Bildsegmenten zu größeren Hierarchieebenen untersucht werden muss. Die Verwendung von irregulären Graphen führt neben der Anpassung an die Bilddaten im Vergleich zu Quadrees in der Regel zu einer signifikanten Verringerung der Anzahl der Segmente [Martinis und Twele, 2010] und somit der durchzuführenden Rechenoperationen bei Parameterschätzung und Inferenz. Des Weiteren kann hier die Anzahl der Hierarchieebenen entsprechend der thematischen Problemstellung sowie den vorliegenden Daten gewählt werden.

Insbesondere bei der rechnergestützten Analyse von hochauflösenden Fernerkundungsdaten gilt es, einen Kompromiss zwischen Automatisierungsgrad und Übertragbarkeit der Methoden zu finden. Im Rahmen von Notfall- und Krisenkartierungen wurde in diesem Zusammenhang jedoch die Erkenntnis gewonnen, dass eine Automatisierung in den meisten Fällen nicht realisierbar ist. Das Konzept einer interaktiven und darüber hinaus allgemein formulierten Methode der Bildanalyse in dieser Arbeit zielt auf die Übertragbarkeit dieser auf unterschiedliche thematische Fragestellungen sowie auf unterschiedliche Sensorprodukte. Sowohl die Kombination von verschiedenen Methoden (Bildmerkmale, Merkmalsselektion, Bildmodelle, Parameterschätzung und Inferenz) als auch die Integration dieser in ein interaktives Rahmenwerk können auf der Grundlage der Ergebnisse in dieser Arbeit als ein sehr wertvoller Ansatz in diesem Zusammenhang gesehen werden. Mit dem vorliegenden Rahmenwerk wird darüber hinaus der Anwendungsbereich von Methoden der statistischen graphenbasierten Bildmodellierung erweitert. Die vorgestellten, mitunter sehr effektiven Inferenzverfahren sowie die gewinnbringende Kombination von Methoden der Bildverarbeitung in einem interaktiven Rahmenwerk verdeutlichen das große Potential des hier vorgeschlagenen Ansatzes.

8 Anhang: Verwendete Bildmerkmale

Statistische Merkmale erster Ordnung

Das Histogramm $h(i)$ eines Segments ist eine Zusammenfassung seiner statistischen Eigenschaften, wobei keine räumlichen Informationen über die Verteilung vorliegen. Teilt man die Werte des Histogramms durch die Anzahl der Pixel N eines Segments, so erhält man die approximierte Wahrscheinlichkeitsdichte $p(i)$ bezüglich des Auftretens von Intensitätswerten $i = 0, \dots, G-1$, wobei G die Anzahl der Grauwerte darstellt. Aus der Wahrscheinlichkeitsdichte lassen sich verschiedene Parameter berechnen, welche die Statistik erster Ordnung quantitativ beschreibt [Meterka und Strzelecki, 1998]:

$$\begin{aligned} \text{Mittelwert: } \mu &= \sum_{i=0}^{G-1} i \cdot p(i) \\ \text{Varianz: } \sigma^2 &= \sum_{i=0}^{G-1} (i - \mu)^2 \cdot p(i) \\ \text{Skewness: } \mu_3 &= \sigma^{-3} \cdot \sum_{i=0}^{G-1} (i - \mu)^3 \cdot p(i) \\ \text{Kurtosis: } \mu_4 &= \sigma^{-4} \cdot \sum_{i=0}^{G-1} (i - \mu)^4 \cdot p(i) - 3 \\ \text{Energie: } E &= \sum_{i=0}^{G-1} p(i)^2 \\ \text{Entropie: } H &= - \sum_{i=0}^{G-1} p(i) \cdot \log_2(p(i)) \end{aligned} \tag{8.1}$$

Die Varianz σ^2 beschreibt die Verteilung der Intensitätswerte um den Mittelwert μ . Die Parameter *skewness* und *kurtosis* sind Indikatoren der Symmetrie beziehungsweise der Spitzigkeit eines Histogramms. Während Mittelwert und Varianz für alle spektralen Kanäle eines Bildes berechnet werden können, werden die übrigen Parameter hier lediglich für den Grauwertkanal bestimmt.

NDVI und weitere Indikatoren

Verhältnisse (Ratios) verbinden mehrere spektrale Kanäle durch arithmetische Operationen. Durch Ratiobildung können etwaige Fehlereinflüsse eliminiert werden, da diese oft eine Korrelation mit verschiedenen spektralen Kanälen aufweisen. Ein sehr bekanntes Verhältnis ist der

normalisierte Differenzen-Vegetationsindex (NDVI). Hierbei werden die Kanäle Rot (R) und nahes Infrarot (NIR) wie folgt kombiniert:

$$NDVI = \frac{NIR - R}{NIR + R}. \quad (8.2)$$

Gesunde Vegetation weist, hervorgerufen durch einen hohen Chlorophyllgehalt, eine niedrige Rückstreuung im roten spektralen Kanal auf. Des Weiteren ist die Rückstreuung im nahen Infrarot hochgradig mit der Vitalität von Vegetation korreliert. Der Index NDVI beschreibt sowohl die Vitalität, als auch die vorkommende Menge von Vegetation. Sein Wertebereich liegt im Intervall $[-1, 1]$. Bei einem Wert von $NDVI = 1$ kann geschlussfolgert werden, dass viel Vegetation vorhanden (NIR groß), und diese zudem noch gesund ist (R klein). Beträgt $NDVI = -1$, so liegt keine bzw. tote Vegetation vor (R groß), was durch eine sehr niedrige Rückstreuung im Kanal NIR bestätigt wird.

Weitere häufig verwendete Merkmale

Die folgenden Merkmale sind ebenfalls für die Charakterisierung eines Segments S_i , $i = 1, \dots, k$ mit k Kanälen und einem mittleren Helligkeitswert von \bar{b} gebräuchlich:

$$\begin{aligned} \text{Maximaler Pixelwert: } & \max(S_i) \\ \text{Minimaler Pixelwert: } & \min(S_i) \\ \text{Maximale Differenz: } & \max_{ij} \left(\frac{|S_i - S_j|}{\bar{b}} \right), \quad i, j = 1, \dots, k \\ \text{Kanalratio: } & \frac{\bar{S}_i}{\bar{b}} \end{aligned} \quad (8.3)$$

Ein Ratio gibt den Anteil eines mittleren Intensitätswertes eines Kanals an der gesamten mittleren Helligkeit \bar{b} wieder.

Farbraumtransformationen

Die Untersuchungen in [Trias-Sanz et al., 2008] ergeben, dass mit einer Anwendung von Texturmerkmalen für die Segmentierung hochauflösender optischer Luftbilder von ländlichen Regionen keine besseren Ergebnisse gegenüber der Verwendung von Farbinformation erzielt werden können. Allerdings wird beobachtet, dass eine Verwendung transformierter Farbkanäle eine bessere Segmentierung produziert, als wenn sie auf Grundlage der Informationen aus den ursprünglichen Farbkanälen (R, G, B) durchgeführt wird. Bei der Anwendung der als optimal identifizierten Farbkanäle auf andere Szenen ist jedoch nicht gewährleistet, dass diese Kombination ebenfalls gute Ergebnisse erzielt. Daher wird empfohlen individuell zu untersuchen, welche Farbkanäle für die jeweilige Anwendung die besten Ergebnisse liefert. Verschiedene Farbraumtransformationen, Texturmerkmale und kombinierte Merkmale aus Farbe und Textur [Hauta-Kasari et al., 1996] werden in [Vansteenkiste et al., 2004] für eine segmentbasierte Klassifikation von hochaufgelösten Satellitenbildern getestet. Auch hier kommen die Autoren zu dem Ergebnis, dass einfache Farbmerkmale in vielen Fällen bessere Ergebnisse im Sinne einer niedrigen Fehlklassifikation liefern. Aufgrund dieser Ergebnisse werden auch in dieser

Arbeit verschiedene Farbraumtransformationen umgesetzt. Diese sind in der Regel sehr einfach zu berechnen und somit auch schnell implementierbar, so dass ein sehr weites Spektrum gängiger Farbräume Anwendung findet.

HSV-Farbraum

Der HSV-Farbraum [Smith, 1978] basiert auf den wahrnehmungsbezogenen Variablen Farbton (*hue*), Sättigung (*saturation*) und Helligkeit (*value*). Der Farbton wird durch einen Winkel definiert, wobei $0^\circ = R$, $120^\circ = G$, $240^\circ = B$ definiert ist. Die Sättigung S bestimmt den Reinheitsgrad und die Helligkeit wird durch den Wert V beschrieben. Die Umrechnung vom RGB-Farbraum (normiert auf das Intervall $[0, 1]$) zum HSV-Farbraum ist durch die folgenden Transformationsgleichungen definiert:

$$\begin{aligned} H &= \begin{cases} 0^\circ, & \max = \min \\ 60 \cdot \left(0 + \frac{G-B}{\max-\min}\right), & \max = R \\ 60 \cdot \left(2 + \frac{B-R}{\max-\min}\right), & \max = G \\ 60 \cdot \left(4 + \frac{R-G}{\max-\min}\right), & \max = B \end{cases} \\ S &= \begin{cases} 0, & \max = 0 \\ \frac{\max-\min}{\max}, & \text{sonst} \end{cases} \\ V &= \max, \end{aligned} \quad (8.4)$$

wobei $\max = \max(R, G, B)$ und $\min = \min(R, G, B)$. Die Werte S und V liegen im Intervall $[0, 1]$ und der Winkel H im Intervall $[0^\circ, 360^\circ]$.

HLS-Farbraum

Mit dem HLS-Farbraum wird eine bessere Annäherung an die Farbwahrnehmung des Menschen erreicht [Luong, 1993]. Ein Punkt in diesem Farbraum wird durch die Variablen H (Farbton), L (Helligkeit) und S (Sättigung) definiert. Auch hier wird der Farbton durch einen Winkel im Intervall $[0^\circ, 360^\circ]$ ausgedrückt, während die Werte L und S wiederum in einem Intervall $[0, 1]$ liegen. Die Umrechnung vom RGB-Farbraum (normiert auf das Intervall $[0, 1]$) zum HLS-Farbraum ist durch die folgenden Transformationsgleichungen definiert:

$$\begin{aligned} L &= \frac{1}{2} (\max + \min) \\ S &= \begin{cases} 0, & \max = \min \\ \frac{\max-\min}{\max+\min}, & L \leq \frac{1}{2} \\ \frac{\max-\min}{2-(\max+\min)}, & L > \frac{1}{2} \end{cases}. \end{aligned} \quad (8.5)$$

H wird entsprechend (8.4) berechnet. Die Werte S und L liegen im Intervall $[0, 1]$ und der Winkel H im Intervall $[0^\circ, 360^\circ]$.

Die Farbräume XYZ, CIE Lab und CIE Luv

Im Jahre 1931 wurden von der CIE (*Commission Internationale de l'Eclairage*), als Alternative zu dem gebräuchlichen Farbtuplet RGB , die drei Primärfarben XYZ definiert [Levkowitz, 1997]. Durch die Kombination positiver Farbkomponenten des RGB -Farbraums sind nicht

alle wahrnehmbaren Farben erzeugbar. Dies kann jedoch mit den künstlichen Grundfarben XYZ erreicht werden. Diese drei Primärfarben definieren einen dreidimensionalen kegelförmigen Farbraum, der alle sichtbaren Farben enthält. Die Ebene $X + Y + Z = 1$ (Schnitt durch den Kegel) definiert alle sichtbaren Farbwerte des CIE-Farbraums. Alle Farben, die den gleichen Farbton jedoch unterschiedliche Helligkeiten aufweisen, werden auf den gleichen Punkt in dieser Ebene abgebildet. Wird diese Ebene in die XY -Ebene projiziert, ergibt sich die Form eines Hufeisens. Die Komponente Z kann durch den Zusammenhang $X + Y + Z = 1$ aus diesem Diagramm bestimmt werden. Y entspricht den Helligkeitsempfindungen des menschlichen Auges und die Werte X und Z beinhalten sämtliche Farbinformation. Die Transformation eines RGB -Farbtriplets in den XYZ -Farbraum kann durch die folgende Vorschrift vollzogen werden:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0,607 & 0,174 & 1,200 \\ 0,299 & 0,587 & 0,114 \\ 0,000 & 0,066 & 1,116 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (8.6)$$

Neben einer schweren Handhabung ist jedoch, wie beispielsweise auch bei den Farbräumen RGB und HSV , keine gleichmäßige Wahrnehmung des menschlichen Auges gewährleistet, d.h. dass die Änderung eines Farbparameters - unabhängig vom Anfangswert - keine äquivalente Änderung in der Farbwahrnehmung bewirkt. In der Wahrnehmung nahezu lineare Modelle sind mit den Farbräumen $CIELab$ und $CIELuv$ gegeben, wobei nichtlineare Zusammenhänge die logarithmische Wahrnehmung des menschlichen Auges nachempfinden sollen [Ford und Roberts, 1998]. Die Transformationsgleichungen vom XYZ -Raum zu den Räumen $CIELab$ und $CIELuv$ sind gegeben mit:

$$\begin{aligned} L &= \begin{cases} 25 \cdot \left[100 \frac{Y}{Y_0} \right]^{1/3} - 16, & \frac{Y}{Y_0} \geq 0,008856 \\ 903,3 \cdot \frac{Y}{Y_0}, & \text{sonst} \end{cases} \\ a &= 500 \cdot \left[\left[\frac{X}{X_0} \right]^{1/3} - \left[\frac{Y}{Y_0} \right]^{1/3} \right] \\ b &= 200 \cdot \left[\left[\frac{Y}{Y_0} \right]^{1/3} - \left[\frac{Z}{Z_0} \right]^{1/3} \right], \end{aligned} \quad (8.7)$$

und

$$\begin{aligned} u &= 13 \cdot L(u' - u_0) \\ v &= 13 \cdot L(v' - v_0), \end{aligned} \quad (8.8)$$

mit

$$\begin{aligned} u' &= \frac{4X}{X + 15Y + 3Z} \\ v' &= \frac{9Y}{X + 15Y + 3Z}, \end{aligned} \quad (8.9)$$

wobei X_0, Y_0, Z_0 das Referenzweiß (z.B. aus $RGB = 255, 255, 255$) definiert und u_0, v_0 durch Ersetzen der Werte XYZ mit diesem Triplet in (8.9) bestimmt werden.

Opponent Farbraum

Der Opponent-Farbraum [Ballard und Brown, 1982] ist durch eine lineare Transformation des *RGB*-Systems definiert:

$$\begin{aligned}wb &= R + G + B \\rg &= R - G \\yb &= 2B - R - G.\end{aligned}\tag{8.10}$$

Durch die Trennung von Farb- und Intensitätsinformation konnte dieser Farbraum erfolgreich bei der Indexierung von Farbbildern [Swain und Ballard, 1991] eingesetzt werden. Die Trennung von Helligkeit und Farbe ermöglicht in diesem Zusammenhang eine gröbere Quantisierung der *wb*-Achse, da diese eine höhere Sensitivität gegenüber Lichtschwankungen aufweist.

Ohta Farbraum

Eine ähnliche Trennung wie beim Opponent-Modell wird auch im Ohta-Farbraum [Ohta et al., 1980] erreicht. Die Transformation ist wie folgt definiert:

$$\begin{aligned}I_1 &= \frac{(R + G + B)}{3} \\I_2 &= \frac{(R - B)}{2} \\I_3 &= \frac{(2G - R - B)}{4}\end{aligned}\tag{8.11}$$

Log-Opponent Farbraum

Die Log-opponent Farbwertkodierung [Faugeras, 1979], [Berens und Finlayson, 2000] hat ebenfalls das Ziel, die nichtlineare menschliche Wahrnehmung zu modellieren und ist gegeben durch:

$$\begin{aligned}l_1 &= \log R - \log G \\l_2 &= \log R + \log G - 2 \log B\end{aligned}\tag{8.12}$$

Unbenannte Transformation

In [Trias-Sanz et al., 2008] wird unter anderem eine Transformation angeführt, ohne einen Namen oder eine Quelle zu nennen. Diese besteht aus den folgenden fünf Komponenten:

$$\begin{aligned}e_{1a} &= \log \frac{R}{R + G + B}, & e_{1b} &= \log \frac{G}{R + G + B} \\e_{2a} &= \frac{R}{G}, & e_{2b} &= \frac{B}{G}, & e_3 &= \log \frac{B}{G}\end{aligned}\tag{8.13}$$

Rot/Cyan-Farbigkeit

Nach [Ohta et al., 1980] sind die Eigenvektoren der Kovarianzmatrix eines Bildes für eine große Anzahl von Bildern mit natürlichem Inhalt annähernd gleich. Die daraufhin in [de Wouwer et al., 1999] definierte Karthunen-Loève-Transformation liefert ein in eine orthogonale Basis

transformiertes Bild (statistisch unkorreliert Farbraumachsen). Die zweite Komponente ist durch

$$k_2 = \frac{R}{2} - \frac{B}{2} \quad (8.14)$$

definiert und wird in den Untersuchungen von [Trias-Sanz et al., 2008] ebenfalls als Merkmal verwendet. Dieses Merkmal entspricht dem Merkmal I_2 der Ohta-Farbraumtransformation(8.11).

P1P2-Farbraum

In [Vertan et al., 2000] wird für die Beschreibung und Klassifikation von Farbbildern eine Familie neuer Farbräume vorgeschlagen. Dazu werden die ursprünglichen Farbwerte zunächst durch

$$r = \frac{R}{R+G+B}, \quad g = \frac{G}{R+G+B}, \quad b = \frac{B}{R+G+B} \quad (8.15)$$

normiert. Da nach diesem Schritt immer $r + g + b = 1$ gilt, reichen lediglich zwei Koordinaten aus, um den rgb -Farbraum zu repräsentieren. Durch eine Rotation des initialen RGB -Systems ergeben sich drei reduzierte Farbrepräsentationen:

$$\begin{aligned} 1) \quad P_1 &= \frac{1}{\sqrt{2}} \frac{G-R}{R+G+B}, \quad P_2 = \frac{1}{\sqrt{6}} \frac{2B-R-G}{R+G+B} \\ 2) \quad P_1 &= \frac{1}{\sqrt{2}} \frac{R-B}{R+G+B}, \quad P_2 = \frac{1}{\sqrt{6}} \frac{2G-R-B}{R+G+B} \\ 3) \quad P_1 &= \frac{1}{\sqrt{2}} \frac{B-G}{R+G+B}, \quad P_2 = \frac{1}{\sqrt{6}} \frac{2R-B-G}{R+G+B}, \end{aligned} \quad (8.16)$$

wobei die am wenigsten signifikanten Farbkanäle des RGB -Farbraumes die Kanäle B (Raum 1), G (Raum 2) und R (Raum 3) darstellen. Aus Gründen des attraktiven Verhältnisses von Komplexität und Laufzeit wird ein *range to average-ratio* verwendet, um den Kanal mit der kleinsten Signifikanz zu identifizieren. Durch die Analyse des Fourier-Energiespektrums eines auf diese Weise transformierten Bildes führte in [Vertan et al., 2000] bei der Texturerkennung zu besseren Ergebnissen, als beispielsweise bei der Verwendung der Ohta-Repräsentation.

l_1, l_2, l_3 -Farbraum

In [Gevers und Smeulders, 1999] werden unter der Annahme weißer Beleuchtung und zweifarbigiger Reflexion neue Farbmodelle für die Erkennung künstlicher dreidimensionaler Objekte in Farbbildern vorgestellt. Der l_1, l_2, l_3 -Farbraum ist definiert durch

$$\begin{aligned} l_1 &= \frac{(R-G)^2}{(R-G)^2 + (R-B)^2 + (G-B)^2} \\ l_2 &= \frac{(R-B)^2}{(R-G)^2 + (R-B)^2 + (G-B)^2} \\ l_3 &= \frac{(G-B)^2}{(R-G)^2 + (R-B)^2 + (G-B)^2}. \end{aligned} \quad (8.17)$$

Unter den oben angesprochenen Annahmen (Shafer's zweifarbiges Reflektionsmodell [Gevers und Smeulders, 1999]) weist dieser Farbraum Invarianz bezüglich Blickwinkel, Objektgeometrie, Schattierung und *highlights* auf.

$h_1h_2h_3$ -Farbraum

Ein neuer Farbraum, welcher Invarianz bezüglich *highlights* im Bild aufweist, wird in [Wesolkowski, 1999] vorgestellt und ist wie folgt definiert:

$$\begin{aligned}h_1 &= R - G \\h_2 &= G - B \\h_3 &= B - R\end{aligned}\tag{8.18}$$

In diesem Farbraum wird durch die Differenzbildung die Helligkeitskomponente eliminiert.

Texturmerkmale

Statistischer Merkmale zweiter Ordnung

Die räumliche Verteilung von Bildpixeln kann mittels Statistiken zweiter Ordnung charakterisiert werden. Hier wird die gemeinsame Verteilung von Pixelpaaren in einer bestimmten Entfernung d und einem Orientierungswinkel θ analysiert. Die Größe der gemeinsamen Verteilungsmatrix (*gray-level co-occurrence-matrix* oder GLCM) ist dabei abhängig von der radiometrischen Auflösung des Bildes. Ein Wert an der Stelle (i, j) der Matrix gibt die Anzahl der vorhandenen Pixelpaare, die den Grauwert i und j in einer Entfernung d und einem Winkel θ aufweisen wieder. Die Rechenzeit und der Speicherbedarf können durch die Reduktion der Intensitätswerte verringert werden, wodurch jedoch ein gewisser Verlust von Texturinformation bewirkt wird. Eine reduzierte Anzahl von skalaren Merkmalen zur Beschreibung von Textur kann direkt aus den Elementen der GLCM berechnet werden [Haralick, 1979].

In [Ruiz et al., 2004] werden verschiedene Texturmerkmale angewendet, um unterschiedliche ländliche und städtische Gebiete zu klassifizieren und eine optimale Kombination von Merkmalen für jeden Landbedeckungstyp zu bestimmen. Die Kombination von GLCM-Texturmerkmalen und spektralen Informationen verbessert hier die Genauigkeit der Klassifikation. Eine Kombination von GLCM-Merkmalen mit beispielsweise Wavelets brachte eine weitere erhebliche Verbesserung der Klassifikationsgenauigkeit. Nach [Trias-Sanz et al., 2008] ist jedoch festzuhalten, dass sich die GLCM-Merkmale nicht für die Beschreibung ländlicher Gebiete in hochauflösenden Bildern eignen. Hier treten periodische Texturen - wenn überhaupt - in großen Intervallen auf, wobei die Orientierung irrelevant, jedoch die Periode selbst sehr diskriminativ ist.

Summen- und Differenzhistogramme werden in [Unser, 1986] als eine Alternative zu den GLCM vorgestellt. Es wird gezeigt, dass mit den Informationen aus einer gemeinsamen Verwendung dieser Histogramme sehr ähnliche Ergebnisse wie mit den GLCM bei Verringerung von Rechenzeit und Speicheranforderungen erreicht werden. Die normalisierte Summe s und Differenz d in einer bestimmten Verschiebung d_1, d_2 ist jeweils definiert als

$$\begin{aligned}s_{k,l} &= y_{k,l} + y_{k+d_1, l+d_2} \\d_{k,l} &= y_{k,l} - y_{k+d_1, l+d_2},\end{aligned}\tag{8.19}$$

wobei y den Grauwert, und (k, l) die Koordinate eines Pixels beschreibt. Die Summen- und Differenzhistogramme über ein Bild D sind den GLCM sehr ähnlich:

$$\begin{aligned}h_s(i, (d_1, d_2)) &= \text{Card} \{(k, l) \in D, s_{k,l} = i\} \\h_d(j, (d_1, d_2)) &= \text{Card} \{(k, l) \in D, d_{k,l} = j\}.\end{aligned}\tag{8.20}$$

Eine Normalisierung erfolgt mit der Teilung der Histogramme durch die Anzahl N der Differenzen bzw. Summen und ergibt die Schätzung der Wahrscheinlichkeitsdichten $P_s(i)$ und $P_d(i)$ der Summen beziehungsweise Differenzen. Von den 14 Merkmalen, welche in [Haralick et al., 1973] vorgestellt werden, können neun Merkmale exakt, und fünf Merkmale approximativ aus den hier beschriebenen Histogrammen berechnet werden. Es folgt eine Auflistung der exakt berechenbaren Merkmale.

$$\begin{aligned}\text{Mittelwert: } f_1 = \mu &= \frac{1}{2} \sum_i i \cdot P_s(i) \\ \text{Varianz: } f_2 = \sigma^2 &= \frac{1}{2} \left(\sum_i (i - 2\mu)^2 \cdot P_s(i) + \sum_j j^2 \cdot P_d(j) \right) \\ \text{Energie: } f_3 &= \sum_i P_s(i)^2 \cdot \sum_j P_d(j)^2 \\ \text{Korrelation: } f_4 &= \frac{1}{2} \left(\sum_i (i - 2\mu)^2 \cdot P_s(i) - \sum_j j^2 \cdot P_d(j) \right) \\ \text{Entropie: } f_5 &= - \sum_i P_s(i) \cdot \log(P_s(i)) - \sum_j P_d(j) \log(P_d(j)) \\ \text{Kontrast: } f_6 &= \sum_j j^2 \cdot P_d(j) \\ \text{Homogenität: } f_7 &= \sum_j \frac{1}{1 + j^2} \cdot P_d(j) \\ \text{Cluster Farbtiefe: } f_8 &= \sum_i (i - 2\mu)^3 \cdot P_s(i) \\ \text{Cluster Prominenz: } f_9 &= \sum_i (i - 2\mu)^4 \cdot P_s(i)\end{aligned}\tag{8.21}$$

Statt der $G \times G$ Elemente in einer GLCM, beinhalten Summen- bzw. Differenzenhistogramme $G \times 2$ Elemente (mit $G = \text{Anzahl der Intensitätsstufen}$). Gegenüber der Verwendung von GLCM sind bei der Berechnung der skalaren Größen keine Doppelsummen zu berechnen. Die Laufzeit sowie der benötigte Speicherplatz für die Berechnung der Kenngrößen werden damit um einen Faktor $G \times 4$ verringert. Des Weiteren kann auch hier eine Quantisierung der Intensitätswerte erfolgen.

Strukturkomplexität

In [Baillard, 1997] wird ein Merkmal vorgestellt, welches Texturen in Abhängigkeit der vorherrschenden Struktur unterscheidet. Dazu wird die Entropie des Histogramms der Gradientenrichtungen herangezogen. In einem ersten Schritt wird der Gradient des Eingangsbildes (respektive Segments) berechnet (Betrag $M(z)$ und Richtung $\theta(z)$ beispielsweise durch einen Sobel-Filter). Die Werte N und N_v beschreiben die Anzahl der Pixel des betrachteten Segments und die Anzahl der Pixel, dessen Betrag des Gradienten höher als ein Schwellwert v ist. Anschließend wird das Histogramm $H(i)$ von den Werten der Richtungen $\theta(z)$ der zuletzt

genannten Pixelmenge erstellt und durch den Wert N_v geteilt. Die Histogramm-Entropie wird anschließend mit

$$E = - \sum_i H(i) \cdot \log_2 H(i) \quad (8.22)$$

berechnet. In [Trias-Sanz et al., 2008] wird als zusätzliches Merkmal das Verhältnis N_v/N verwendet. Es werden verschiedene Farbkanäle und Texturmerkmale für die Segmentierung hochauflösender Luftbilder ländlicher Gebiete in verschiedenen Kombinationen getestet. Anhand der Ergebnisse kann geschlussfolgert werden, dass Texturmerkmale die Segmentierung dieses Bildtyps nicht verbessern. Neben verschiedenen Farbraumtransformationen, welche die Qualität der Segmentierung gegenüber der Verwendung der ursprünglichen spektralen Kanäle verbessern, werden lediglich mit der hier vorgestellten Strukturkomplexität gute Ergebnisse erzielt.

Texturinformation aus lokalen Histogrammen

In [Lowitz, 1983] wird die Berechnung zweier Merkmale (*module* und *state*) eines jeden Bildpunktes aus lokalen Histogrammen vorgestellt. Die daraus resultierenden Bilder beinhalten lokale räumliche Informationen, und sind somit eng mit der Textur verknüpft. Grundlage dieser beiden Merkmale ist die auf dem Gebiet der Quantenphysik gebräuchlichen Informationsmessung bezüglich eines beobachteten Wertes a und der Messung der Unschärfe Δa nach [Mackay, 1955]:

$$I_M(a) = \frac{\bar{a}}{\Delta a}. \quad (8.23)$$

Bevor das lokale Histogramm bekannt ist, kann als eine Anfangsschätzung ein gleich verteiltes, also flaches Histogramm h_0 herangezogen werden. Der Vergleich eines lokalen Histogramms mit diesem Basishistogramm kann mit einer Distanzmessung zwischen diesen beiden Histogrammen erfolgen und liefert eine Informationsmessung beziehungsweise Entropie zwischen den Zuständen vor und nach der Bestimmung des lokalen Histogramms:

$$I_{MH}(h, h_0) = \sum_{i=1}^r \frac{|n_i - N/r|}{\sqrt{n_i(1 - n_i/N) + N/r(1 - 1/r)}}, \quad (8.24)$$

wobei N die Anzahl der Pixel, r die Anzahl der Histogrammstufen und n_i die Anzahl der Pixel pro Stufe i ist. Die so bestimmte Entropie eines Histogramms wird in [Lowitz, 1983] *module* genannt. Der Kennwert *state* ist der erste Index des Intensitätslevels, welches mit dem Maximum des lokalen Histogramms korrespondiert. Bei den Untersuchungen in [Lowitz, 1983] wurden die *module*- und *state*-Bilder von Landsat-TM und Daedalus-Szenen (ähnliche Auflösung wie Landsat), sowie von Portraitfotos berechnet, wobei zuvor eine Quantisierung auf $r = 16$ Intensitätswerte erfolgte. Bei einer Fenstergröße von $N = 3 \times 3$ ist das Bild der *module*-Werte ein Konturbild, welches nach einer Schwellwertbildung mit dem Ergebnis einer Sobel-Filterung vergleichbar ist. Für $N = 5 \times 5$ bis $N = 9 \times 9$ werden die Konturen verwischt und das *module*-Bild repräsentiert etwas verrauschte Texturinformation. Bei einer Fenstergröße von $N = 16 \times 16$ ist im Ergebnis keine Information über die Konturen enthalten, vielmehr ist sie sehr texturähnlich.

Literaturverzeichnis

- [Aksoy und Haralick, 2000] Aksoy, S. und Haralick, R. M. (2000). Feature normalization and likelihood-based similarity measures for image retrieval. *Pattern Recognition Letters*, 22:563–582.
- [Annoni et al., 2010] Annoni, A., Craglia, M., de Roo, A. und San-Miguel, J. (2010). *Earth Observations and Dynamic Mapping: Key Assets for Risk Management*, Seiten 3–21. Geographic Information and Cartography for Risk and Crisis Management. Springer, Berlin Heidelberg.
- [Atkinson und Lewis, 2000] Atkinson, P. M. und Lewis, P. (2000). Geostatistical classification for remote sensing: An introduction. *Computers and Geosciences*, 26:361–371.
- [Awasthi et al., 2007] Awasthi, P., Gagrani, A. und Ravindran, B. (2007). Image modeling using tree structured conditional random fields. In *International Joint Conferences on Artificial Intelligence*.
- [Baatz und Schäpe, 2000] Baatz, M. und Schäpe, A. (2000). Multiresolution segmentation - An optimization approach for high quality multi-scale image segmentation. In Strobl, J., Blaschke, T. und Griesebner, G., editors, *Angewandte Geographische Informationsverarbeitung XII*, Seiten 12–23. Wichmann-Verlag, Heidelberg.
- [Baillard, 1997] Baillard, C. (1997). *Analyse d’Images Aéienne Stéréo Pour la Restitution 3-d en Milieu Urbain*. PhD thesis, ENST, Paris, France.
- [Ballard und Brown, 1982] Ballard, D. H. und Brown, C. M. (1982). *Computer Vision*. Prentice Hall: New York.
- [Berens und Finlayson, 2000] Berens, J. und Finlayson, G. D. (2000). Log-opponent chromaticity coding of color space. In *Proceedings International Conference on Pattern Recognition, IAPR*, volume 1, Seiten 206–211.
- [Besag, 1974] Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B*, 36:192–236.
- [Besag, 1975] Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician*, 24(3):179–195.
- [Besag, 1977] Besag, J. (1977). Efficiency of pseudo-likelihood estimation of simple Gaussian fields. *Biometrika*, 64:616–618.
- [Besag, 1986] Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, Series B*, 48:259–302.
- [Bishop, 2006] Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer New York, 1st edition.

- [Blaschke, 2010] Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65:2–16.
- [Blaschke und Strobl, 2001] Blaschke, T. und Strobl, J. (2001). What’s wrong with pixels? Some recent developments interfacing remote sensing and GIS. *GIS*, 14(6).
- [Bouman und Shapiro, 1994] Bouman, C. A. und Shapiro, M. (1994). A multiscale random field model for Bayesian image segmentation. *IEEE Transactions on Image Processing*, 3(2):162–177.
- [Bow, 1992] Bow, S. T. (1992). *Pattern Recognition and Image Preprocessing*. Marcel Dekker Inc., New York, NY.
- [Bruzzone und Prieto, 2000] Bruzzone, L. und Prieto, D. F. (2000). Automatic analysis of the difference image for unsupervised change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 38(3):1171–1182.
- [Cadez und Smyth, 1998] Cadez, I. V. und Smyth, P. (1998). Parameter estimation for inhomogeneous Markov random fields using pseudo likelihood. In *IEEE Proc. Conf. on Computer Vision and Pattern Recognition (CVPR)*, Seiten 262–269.
- [Cai et al., 2007] Cai, W. L., Chen, S. C. und Zhang, D. Q. (2007). Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation. *Pattern Recognition*, 40:825–838.
- [Chellappa und Jain, 1993] Chellappa, R. und Jain, A. (1993). *Markov Random Fields: Theory and Applications*. Academic Press.
- [Cheng und Bouman, 2001] Cheng, H. und Bouman, C. A. (2001). Multiscale Bayesian segmentation using a trainable context model. *IEEE Trans. on Image Processing*, 10:511–525.
- [Cheng et al., 1997] Cheng, J., Bell, D. A. und Liu, W. (1997). Learning belief networks from data: an information theory based approach. In *Proceedings of the Sixth ACM International Conference on Information and Knowledge Management*, Seiten 325–331. New York: ACM Press.
- [Collet und Murtagh, 2004] Collet, C. und Murtagh, F. (2004). Multiband segmentation based on a hierarchical Markov model. *Pattern Recognition*, 37:2337–2347.
- [Constantinopoulos et al., 2006] Constantinopoulos, C., Titsias, M. K. und Likas, A. (2006). Bayesian feature and model selection for gaussian mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28:1013–1018.
- [Cross und Jain, 1983] Cross, G. C. und Jain, A. K. (1983). Markov random field texture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5:25–39.
- [Dai et al., 2009] Dai, D., Yang, W. und Triggs, B. (2009). SAR image labeling with hierarchical Markov aspect models. *IEEE Transactions on Geoscience and Remote Sensing*.
- [Dash und Liu, 1997] Dash, M. und Liu, H. (1997). Feature selection for classification. *Intelligent Data Analysis*, 1:131–156.

- [De Smet und Pries, 2000] De Smet, P. und Pries, R. L. (2000). Implementation and analysis of an optimized rainfalling watershed algorithm. In *SPIE/EI, IS&T/SPIE's 12th Annual Symposium Electronic Imaging 2000: Science and Technology*, Conference: Image and Video Communications and Processing (ei26), San Jose, California USA.
- [de Wouwer et al., 1999] de Wouwer, G. V., Scheunders, P., Livens, S. und Dyck, D. V. (1999). Wavelet correlation signatures for color texture characterization. *Pattern Recognition*, 32(3):443–451.
- [Dempster et al., 1977] Dempster, A. P., Laird, N. M. und Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM-algorithm. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, 39(1):1–38.
- [Derrode und Pieczynski, 2000] Derrode, S. und Pieczynski, W. (2000). Signal and image segmentation using pairwise Markov chains. *IEEE Trans. Signal Processing*, 59(9):2477–2489.
- [Dietterich et al., 2004] Dietterich, T. G., Ashenfelder, A. und Bulatov, Y. (2004). Training conditional random fields via gradient tree boosting. In *In Proceedings of the 21th International Conference on Machine Learning (ICML)*, Seiten 217–224. ACM.
- [Faugeras, 1979] Faugeras, O. (1979). Digital color image processing within the framework of a human visual model. *IEEE Transaction on Acoustics, Speech and Signal Processing*, 27(4):380–393.
- [Feng et al., 2002] Feng, X., Williams, C. K. und Felderhof, S. N. (2002). Combining belief networks and neural networks for scene segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(4):467–483.
- [Ferri et al., 1994] Ferri, F., Pudil, P., Hatef, M. und Kittler, J. (1994). Comparative study of techniques for large scale feature selection. In Gelsema, E. und Kanal, L., editors, *Pattern Recognition in Practice IV*, Seiten 403–413.
- [Ford und Roberts, 1998] Ford, A. und Roberts, A. (1998). Colour space conversions.
- [Forney, 1973] Forney, G. D. (1973). The viterbi algorithm. *IEEE Proceedings*, 61:268–278.
- [Frey und Butenuth, 2010] Frey, D. und Butenuth, M. (2010). Assessment system of GIS-objects using multi-temporal imagery for near-realtime disaster management. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Comm. VII, Part A*.
- [Gagrani, 2002] Gagrani, A. (2002). Image modeling using hierarchical conditional random field. Master's thesis, Department of Computer Science and Engineering, Indian Institute of Technology Madras.
- [Geman und Geman, 1984] Geman, S. und Geman, D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.
- [Gevers und Smeulders, 1999] Gevers, T. und Smeulders, A. W. M. (1999). Color-based object recognition. *Pattern Recognition*, 32:453–464.

- [Gidas, 1989] Gidas, B. (1989). A renormalization group approach to image processing problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:164–180.
- [GRASS Team, 2011] GRASS Team (2011). http://grass.osgeo.org/wiki/Main_Page.
- [Guyon, 2003] Guyon, I. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182.
- [Haralick, 1979] Haralick, R. M. (1979). Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804.
- [Haralick et al., 1973] Haralick, R. M., Shanmugan, K. und Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3:610–621.
- [Hauta-Kasari et al., 1996] Hauta-Kasari, M., Pakkinen, J., Jaaskelainen, T. und Lenz, R. (1996). Generalized cooccurrence matrix for multispectral texture analysis. In *Proceedings of the 13th International Conference on Pattern Recognition*, volume 2, Seiten 785–789.
- [Hay et al., 2003] Hay, G. J., Blaschke, T., Marceau, D. J. und Bouchard, A. (2003). A comparison of three image-object methods for the multiscale analysis of landscape structure. *ISPRS Journal of Photogrammetry & Remote Sensing*, 1253:1–19.
- [He et al., 2006] He, X., Zemel, R. S. und Ray, D. (2006). Learning and incorporating top-down cues in image segmentation. In *In ECCV*, Seiten 338–351. Springer.
- [Hegarat-Masclé und Vidal-Madjar, 1996] Hegarat-Masclé, S. L. und Vidal-Madjar, D. (1996). Applications of simulated annealing to SAR image clustering and classification problems. *International Journal of Remote Sensing*, 17(9):1761–1776.
- [Hese, 2001] Hese, S. M. (2001). *Entwicklung und Anwendung geostatistischer Verfahren zur strukturellen Klassifizierung von Forstbeständen in hochauflösenden Daten der HRSC-A*. PhD dissertation, Technische Universität Berlin.
- [Hinton, 2002] Hinton, J. (2002). Training products of experts by minimizing contrastive divergence. *Neural Computation*, 14(6):1771–1800.
- [Hoberg und Rottensteiner, 2010] Hoberg, T. und Rottensteiner, F. (2010). Classification of settlement areas in remote sensing imagery using conditional random fields. In Wagner, W. und Székely, B., editors, *ISPRS TC VII Symposium - 100 Years ISPRS*, volume XXXVIII, Part 7A, Vienna, Austria. IAPRS.
- [Hubert-Moy et al., 2001] Hubert-Moy, L., Cotonnec, A., Le Du, L., Chardin, A. und Pérez, P. (2001). A comparison of parametric classification procedures of remotely sensed data applied on different landscape units. *Remote Sensing of Environment*, 75:147–187.
- [Irving et al., 1997] Irving, W. W., Fieguth, P. W. und Willsky, A. S. (1997). An overlapping tree approach to multiscale stochastic modeling and estimation. *IEEE Trans. on Image Processing*, 6:1517–1529.
- [Jain et al., 2000] Jain, A. K., Duin, R. P. W. und Mao, J. (2000). Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37.

- [John et al., 1994] John, G. H., Kohavi, R. und Pfleger, K. (1994). Irrelevant features and the subset selection problem. In *International Conference on Machine Learning, Journal version in AIJ*, Seiten 121–129.
- [Kartikeyan et al., 1994] Kartikeyan, B., Gopalakrishna, B., Kalubarme, M. H. und Majumder, K. L. (1994). Contextual techniques for classification of high and low resolution remote sensing data. *International Journal of Remote Sensing*, 15(5):1037–1051.
- [Kersten und Gähler, 2010] Kersten, J. und Gähler, M. (2010). A framework for satellite image classification in the context of crisis mapping using Markov random fields. In *Proceedings of RSPSoc 2010 Annual Conference*.
- [Kersten et al., 2010] Kersten, J., Gähler, M. und Voigt, S. (2010). A general framework for fast and interactive classification of optical vhr satellite imagery using hierarchical and planar Markov random fields. *PFG Photogrammetrie, Fernerkundung, Geoinformation*, 6:439–449.
- [Kolaczyk et al., 2005] Kolaczyk, E. D., Ju, J. und Gopal, S. (2005). Multiscale, multigranular statistical image segmentation. *Journal of the American Statistical Association*, 100:1358–1369.
- [Köppen, 2000] Köppen, M. (2000). The curse of dimensionality. In *5th Online World Conference on Soft Computing in Industrial Applications (WSC5)*.
- [Kudo und Sklansky, 2000] Kudo, M. und Sklansky, J. (2000). Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition*, 33:25–41.
- [Kumar, 2005] Kumar, S. (2005). *Models for Learning Spatial Interactions in Natural Images for Context-Based Classification*. Dissertation, Carnegie Mellon University, The Robotics Institute, School of Computer Science.
- [Kumar und Hebert, 2003a] Kumar, S. und Hebert, M. (2003a). Discriminative random fields: A discriminative framework for contextual interaction in classification. In *International Conference on Computer Vision*, Seiten 1150–1157.
- [Kumar und Hebert, 2003b] Kumar, S. und Hebert, M. (2003b). Man-made structure detection in natural images using a causal multiscale random field. In *In Proc. IEEE Int. Conf. on Comp. Vision and Pattern Recog*, Seiten 119–126. CVPR.
- [Kux und Pinho, 2006] Kux, H. J. H. und Pinho, C. M. D. (2006). Object-oriented analysis of high-resolution satellite images for intra-urban land cover classification: Case study in São José Dos. In *International Conference on Object-Based Image Analysis*.
- [Laferté et al., 2000] Laferté, J.-M., Pérez, P. und Heitz, F. (2000). Discrete Markov image modeling and inference on the quadtree. *IEEE Transactions on Image Processing*, 9(3):390–404.
- [Lafferty et al., 2001] Lafferty, J., McCallum, A. und Pereira, F. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the 18th International Conference on Machine Learning*, Seiten 282–289. Morgan Kaufmann, San Francisco, CA.

- [Lang und Blaschke, 2003] Lang, S. und Blaschke, T. (2003). Hierarchical object representation - comparative multi-scale mapping of anthropogenic and natural features. *International Archives of Photogrammetry, Remote Sensing and spatial information sciences.*, „XXXIV-3/W8“:181–186.
- [Langley, 1994] Langley, P. (1994). Selection of relevant features in machine learning. In *In Proceedings of the AAAI Fall Symposium on Relevance*, Seiten 1–5.
- [Lasserre und Bishop, 2007] Lasserre, J. und Bishop, C. M. (2007). Generative or discriminative? Getting the best of both worlds. *Bayesian Statistics*, 8:3–24.
- [Levkowitz, 1997] Levkowitz, H. (1997). *Color Theory and Modeling for Computer Graphics, Visualization and Multimedia Applications*. Kluwer Academic Publishers: Norwell MA, USA.
- [Li et al., 2000] Li, J., Najmi, A. und Gray, R. M. (2000). Image classification by a two dimensional hidden Markov model. *IEEE Trans. Signal Processing*, 48.
- [Li, 2009] Li, S. Z. (2009). *Markov Random Field Modeling in Image Analysis*. Springer Publishing Company, Incorporated, 3rd edition.
- [Li et al., 2009] Li, Y., Dong, M. und Hua, J. (2009). Simultaneous localized feature selection and model detection for gaussian mixtures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31:953–960.
- [Liu et al., 2008] Liu, D., Song, K., Townshend, J. R. und Gong, P. (2008). Using local transition probability models in Markov random fields for forest change detection. *Remote Sensing of Environment*, 112(5):2222–2231. Earth Observations for Terrestrial Biodiversity and Ecosystems Special Issue.
- [Lowitz, 1983] Lowitz, G. E. (1983). Can a local histogram really map texture information? *Pattern Recognition*, 16:141–147.
- [Lu und Weng, 2007] Lu, D. und Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5):823–870.
- [Lucchese und Mitra, 2001] Lucchese, L. und Mitra, S. (2001). Color image segmentation: A state-of-the-art survey. In *Proceedings of the Indian National Science Academy*, volume 2.
- [Luetten et al., 1993] Luetten, M. R., Karl, W. C. und Willsky, A. S. (1993). Efficient multiscale regularization with applications to the computation of optical flow. *IEEE Transactions on image processing*, 3:41–64.
- [Luong, 1993] Luong, Q. T. (1993). *Color in Computer Vision*, Seiten 311–368. World Scientific Publishing Company.
- [Mackay, 1955] Mackay, M. D. (1955). Quantal aspects of scientific information. *Trans. Inf. Theory*, 1.
- [Magnussen et al., 2004] Magnussen, S., Boudewyn, P. und Wulder, M. (2004). Contextual classification of landsat TM images to forest inventory cover types. *International Journal of Remote Sensing*, 25:2421–2440.

- [Marceau et al., 1994] Marceau, D. J., Howarth, P. J. und Gratton, D. J. (1994). Remote sensing and the measurement of geographical entities in a forest environment 1: The scale and spatial aggregation problem. *Remote Sensing of Environment*, 49:93–104.
- [Marida et al., 1979] Marida, K., Kent, J. und Bibby, J. (1979). *Multivariate Analysis*. London: Academic Press.
- [Marill und Green, 1963] Marill, T. und Green, D. M. (1963). On the effectiveness of receptors in recognition system. *IEEE Trans. Inf. Theory*, 9:11–17.
- [Markov, 1913] Markov, A. A. (1913). An example of statistical investigation in the text of 'eugene onyegin' illustrating a couple of 'tests' in chains. *Proc. Aca. Sci. St. Petersburg*, IV(7):153.
- [Marroquin et al., 1987] Marroquin, J., Mitter, S. und Poggio, T. (1987). Probabilistic solution of ill-posed problems in computational vision. Technical report, Massachusetts Institute of Technology, Cambridge, MA, USA.
- [Martinis und Twele, 2010] Martinis, S. und Twele, A. (2010). A hierarchical spatio-temporal Markov model for improved flood mapping using multi-temporal X-band SAR data. *Remote Sensing*, 2(9):2240–2258.
- [Martinis et al., 2011] Martinis, S., Twele, A. und Voigt, S. (2011). Unsupervised extraction of flood-induced backscatter changes in SAR data using Markov image modeling on irregular graphs. *IEEE Transactions on Geoscience and Remote Sensing*, (in print), 48(12).
- [Mather, 2004] Mather, P. M. (2004). *Computer Processing of Remotely-Sensed Images: An Introduction*. John Wiley & Sons Ltd.
- [Meinel et al., 2001] Meinel, G., Neubert, M. und Reder, J. (2001). Pixelorientierte versus segmentorientierte Klassifikation von IKONOS-Satellitenbilddaten - ein Methodenvergleich. *Zeitschrift für Photogrammetrie, Fernerkundung und Geoinformation*, Seiten 157–170.
- [Meterka und Strzelecki, 1998] Meterka, A. und Strzelecki, M. (1998). Texture analysis methods - a review. Technical report, Institute of Electronics, Technical University of Lodz.
- [Murphy, 2002] Murphy, K. P. (2002). *Dynamic Bayesian networks: representation, inference and learning*. PhD thesis, University of California, Berkeley.
- [Myint et al., 2007] Myint, S. W., Wentz, E. A. und Purkis, S. J. (2007). Employing spatial metrics in urban land-use/land-cover mapping: Comparing the Getis and Geary indices. *Photogrammetric Engineering and Remote Sensing*, 73:1403–1415.
- [Narendra und Fukunaga, 1977] Narendra, P. M. und Fukunaga, K. (1977). A branch and bound algorithm for feature subset selection. *IEEE Trans. Computers*, 26:917–922.
- [Neal und Hinton, 1998] Neal, R. und Hinton, G. E. (1998). A view of the EM-algorithm that justifies incremental, sparse, and other variants. In *Learning in Graphical Models*, Seiten 355–368. Kluwer Academic Publishers.

- [Neubert, 2005] Neubert, M. (2005). *Bewertung, Verarbeitung und segmentbasierte Auswertung sehr hoch auflösender Satellitendaten vor dem Hintergrund landschaftsplanerischer und landschaftsökologischer Anwendungen*. Dissertation, Technische Universität Dresden, Fakultät Forst-, Geo- und Hydrowissenschaften.
- [Nussbaum et al., 2006] Nussbaum, S., Niemeyer, I. und Canty, M. J. (2006). SEaTH - A new tool for automated feature extraction in the context of object-based image analysis. In *Proceedings of the 1st International Conference on Object-based Image Analysis (OBIA 2006)*, ISPRS Volume No. XXXVI - 4/C42.
- [Ohta et al., 1980] Ohta, Y. I., Kanade, T. und Sakai, T. (1980). Color information for region segmentation. *Computer Vision, Graphics and Image Processing*, 13:222–241.
- [Ouyang et al., 2006] Ouyang, Y., Ma, J. und Dai, Q. (2006). Bayesian multinet classifier for classification of remote sensing data. *International Journal of Remote Sensing*, 27:4943–4961.
- [Pal und Pal, 1993] Pal, N. R. und Pal, S. K. (1993). A review on image segmentation techniques. *Pattern Recognition*, 26(9):1277–1294.
- [Pavlidis, 1986] Pavlidis, T. (1986). A critical survey of image analysis methods. In *International Conference on Pattern Recognition*, Seiten 502–511.
- [Pearl, 1988] Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.
- [Peleg et al., 1984] Peleg, S., Naor, J., Hartley, R. und Avnir, D. (1984). Multiple resolution texture analysis and classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(4):518–523.
- [Pérez et al., 2000] Pérez, P., Chardin, A. und Laferté, J.-M. (2000). Noniterative manipulation of discrete energy-based models for image analysis. *Pattern Recognition*, 33(4):573–586.
- [Pudil et al., 1994] Pudil, P., Novovicova, J. und Kittler, J. (1994). Floating search methods in feature selection. *Pattern Recognition Letters*, 15(11):1119–1125.
- [Reunanen, 2003] Reunanen, J. (2003). Overfitting in making comparisons between variable selection methods. *Journal of Machine Learning Research*, 3:1371–1382.
- [Richard et al., 2004] Richard, X. H., Zemel, R. S. und Carreira-Perpiñán, M. . (2004). Multiscale conditional random fields for image labeling. In *In CVPR*, Seiten 695–702.
- [Ronen et al., 1995] Ronen, O., Rohlicek, J. R. und Ostendorf, M. (1995). Parameter estimation of dependence tree models using the EM-algorithm. *IEEE Signal Processing Letters*, 2(8):157–159.
- [Ruiz et al., 2004] Ruiz, L. A., Fdez-Sarría, A. und Recio, J. A. (2004). Texture feature extraction for classification of remote sensing data using wavelet decomposition: A comparative study. In *International Archives of Photogrammetry and Remote Sensing. Vol. XXXV, ISSN*, Seiten 1682–1750.

- [Schneider et al., 2000] Schneider, T., de Kok, R., Buck, A. und Manakos, I. (2000). Objekt-orientierte Bildanalyse - Paradigmenwechsel in der thematischen Auswertung von Erdbeobachtungen? In für Landnutzungsplanung und Naturschutz, L., editor, *Aktuelle Forschungsberichte, Festschrift zur Emeritierung von Prof. Dr. U. Ammer*, Seiten 234–258, Freising, Germany. TUM.
- [Schneiderhan et al., 2010] Schneiderhan, T., Gaehler, M., Kranz, O. und Voigt, S. (2010). Insights into the emergency mapping service within the GMES project SAFER - Highlights, main achievements and challenges. In *Proceedings of the ESA Living Planet Symposium*, Bergen, Norway.
- [Sharma und Sarkar, 1998] Sharma, K. M. und Sarkar, A. (1998). A modified contextual classification technique for remote sensing data. *Photogrammetric Engineering and Remote Sensing*, 64:273–280.
- [Siedlecki und Sklansky, 1989] Siedlecki, W. und Sklansky, J. (1989). A note on genetic algorithm for large-scale feature selection. *Pattern Recognition Letters*, 10(5):335–347.
- [Smith, 1978] Smith, A. R. (1978). Color gamut transform pairs. *Association for Computing Machinery's Special Interest Group on Computer Graphics and Interactive Techniques*, 78:12–19.
- [Somol et al., 1999] Somol, P., Pudil, P., Novovicová, J. und Paclík, P. (1999). Adaptive floating search methods in feature selection. *Pattern Recognition Letters*, 20:1157–1163.
- [Sun et al., 2010] Sun, X., Wang, H. und Fu, K. (2010). Automatic detection of geospatial objects using taxonomic semantics. *IEEE Geoscience and Remote Sensing Letters*, 7(1):23–27.
- [Swain und Ballard, 1991] Swain, M. J. und Ballard, D. H. (1991). Color indexing. *International Journal of Computer Vision*, 7(1):11–32.
- [Townshend et al., 2000] Townshend, J. R. G., Huang, C., Kalluri, S. N. V., Defries, R. S. und Liang, S. (2000). Beware of per-pixel characterization of land cover. *International Journal of Remote Sensing*, 21(4):829–843.
- [Trias-Sanz et al., 2008] Trias-Sanz, R., Stamon, G. und Louchet, J. (2008). Using colour, texture, and hierarchical segmentation for high-resolution remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(2):156–168.
- [Trimble, 2010] Trimble (2010). *Reference Book, eCognition Developer 8.64.0*. Trimble Germany GmbH.
- [Unser, 1986] Unser, M. (1986). Sum and difference histograms for texture classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(1):118–125.
- [van Laarhoven und Aarts, 1987] van Laarhoven, P. J. M. und Aarts, E. H. L. (1987). *Simulated Annealing: Theory and Applications*. Dordrecht: D. Reidel Publishing Company.
- [Vansteenkiste et al., 2004] Vansteenkiste, E., Schoutteet, A., Gautama, S. und Philips, W. (2004). Comparing color and textural information in very high resolution satellite image

- classification. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2004)*.
- [Vertan et al., 2000] Vertan, C., Boujemaa, N., Rocquancourt, I. und Imedia, P. (2000). Color texture classification by normalized color space representation. In *Proceedings of the International Conference on Pattern Recognition*.
- [Ververidis und Kotropoulos, 2008] Ververidis, D. und Kotropoulos, C. (2008). Gaussian mixture modeling by exploiting the mahalanobis-distance. *IEEE Transactions on Signal Processing*, 56:2797–2811.
- [Voigt et al., 2007] Voigt, S., Kemper, T., Riedlinger, T., Kiefl, R., Scholte, K. und Mehl, H. (2007). Assessment system of GIS-objects using multi-temporal imagery for near-realtime disaster management. *IEEE Transactions on Geoscience and Remote Sensing*, 45(6):1520–1528.
- [Wallach, 2002] Wallach, H. (2002). Efficient training of conditional random fields. Master’s thesis, School of Cognitive Science, University of Edinburgh.
- [Wang, 1990] Wang, F. (1990). Improving remote sensing image analysis through fuzzy information representation. *Photogrammetric Engineering and Remote Sensing*, 8:1163–1169.
- [Wesolkowski, 1999] Wesolkowski, S. B. (1999). Color image edge detection and segmentation: A comparison of the vector angle and the euclidean distance color similarity measures. Master’s thesis, University of Waterloo, Ontario, Canada.
- [Whitney, 1971] Whitney, A. W. (1971). A direct method of nonparametric measurement selection. *IEEE Transactions on Computers*, 20(9):1100–1103.
- [Willsky, 2002] Willsky, A. (2002). Multiresolution Markov models for signal and image processing. *Proceedings of the IEEE*, 90(8):1396–1458.
- [Wilson und Li, 2002] Wilson, R. und Li, C.-T. (2002). A class of discrete multiresolution random fields and its application to image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(1):42–56.
- [Yang et al., 2009] Yang, W., Dai, D., Triggs, B. und Xia, G.-S. (2009). Semantic labeling of SAR images with hierarchical Markov aspect models. Report hal-00433600.
- [Zhong et al., 2007] Zhong, P., Liu, F. und Wang, R. (2007). A new MRF framework with dual adaptive contexts for image segmentation. In *International Conference on Computational Intelligence and Security, 2007*, Seiten 351 –355.